GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: May 14, 2000, 18:45:20 ; Search time 871.67 Seconds
(without alignments)
5114.947 Million cell updates/sec

Title: US-09-223-796-1
Perfect score: 1100
Sequence: 1 gcacgagccacagccagcta.....attaaaaaaaaaaaaaaaaa 1100

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: gb_gss1:*
83: gb_gss2:*
84: gb_gss3:*
85: gb_gss4:*
86: em_gss1:*
87: em_gss2:*
88: em_gss3:*
89: em_gss4:*
90: gb_gss5:*
91: gb_gss6:*
92: gb_gss7:*
93: gb_gss8:*
94: gb_gss9:*
95: em_gss5:*
96: em_gss6:*
97: em_gss7:*
98: em_gss8:*
99: em_gss9:*
100: em_gss10:*
101: em_gss11:*
102: gb_gss10:*
103: gb_gss11:*
104: em_gss12:*
105: gb_gss12:*
106: gb_gss13:*
107: gb_gss14:*
108: gb_gss15:*
109: gb_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	% Query

	No.	Score	Match	Length	DB	ID	Description
c	1	595.6	54.1	602	43	AI104547	AI104547 EST213836
c	2	570.4	51.9	680	45	AI232721	AI232721 EST229409
c	3	446.6	40.6	462	43	AI233681	AI233681 EST230369
	4	383.4	34.9	456	27	AA034568	AA034568 mi49a11.r
c	5	382.8	34.8	386	42	AI136965	AI136965 UI-R-C2p-
	6	376.6	34.2	482	41	AI048873	AI048873 uc76f02.y
	7	373	33.9	444	27	AA032667	AA032667 mi32b07.r
	8	362	32.9	454	42	AI154070	AI154070 ud76a06.r
c	9	334.4	30.4	345	81	AW434638	AW434638 UI-R-BJ0p
c	10	327.4	29.8	540	94	AQ055548	AQ055548 CIT-HSP-2
	11	307.4	27.9	394	28	AA065442	AA065442 ml38h02.r
	12	302.4	27.5	446	27	W83837	W83837 mf26c06.r1
c	13	296.6	27.0	569	36	C77953	C77953 C77953 Mous
	14	296.4	26.9	374	33	AA450933	AA450933 vf85a11.r
c	15	286.6	26.1	574	36	C78711	C78711 C78711 Mous
	16	285	25.9	350	26	W67032	W67032 me29f10.r1
c	17	276	25.1	501	90	AQ798652	AQ798652 HS_2238_B
	18	260.2	23.7	537	27	AA029831	AA029831 zk10b06.r
c	19	243.4	22.1	548	38	AA780068	AA780068 zj24h04.s
	20	234.4	21.3	260	30	AA239639	AA239639 mv25a05.r
c	21	234.2	21.3	487	45	AI178891	AI178891 EST222573
	22	233	21.2	300	28	AA067980	AA067980 mm56f09.r
c	23	224.2	20.4	520	63	AW005581	AW005581 wz87h11.x
	24	214.4	19.5	337	81	AW418433	AW418433 56258 MAR
c	25	213.8	19.4	525	61	AI871472	AI871472 wl67e01.x
c	26	213.8	19.4	527	64	AW025527	AW025527 wu97d10.x
c	27	204.6	18.6	424	44	AU045080	AU045080 AU045080
c	28	195.2	17.7	569	44	AU043460	AU043460 AU043460
	29	190.6	17.3	464	26	W86176	W86176 zh59g11.r1
c	30	177	16.1	318	48	AI578514	AI578514 UI-R-AA0-
c	31	174.4	15.9	491	26	W93417	W93417 zd95d01.s1
c	32	171.6	15.6	447	39	AA836105	AA836105 od15c01.s
c	33	164.6	15.0	431	43	AI240879	AI240879 qj94e02.x
c	34	164.6	15.0	437	44	AI276262	AI276262 ql65f06.x
c	35	158.6	14.4	413	42	AI076231	AI076231 oy65c09.x
c	36	157.6	14.3	411	79	AW261921	AW261921 xp87f11.x
c	37	157.6	14.3	418	43	AI189612	AI189612 qd32g06.x
	38	155.6	14.1	330	32	AA379908	AA379908 EST92801
	39	154.4	14.0	361	26	W92672	W92672 zh79d09.r1
c	40	152.6	13.9	406	34	AA464962	AA464962 zx80f06.s
c	41	150.2	13.7	410	39	AA857959	AA857959 oe33g04.s
c	42	148.4	13.5	398	43	AI200129	AI200129 qf90h05.x
	43	148.4	13.5	466	21	R20463	R20463 yg33h09.r1
c	44	146.6	13.3	418	26	W86177	W86177 zh59g11.s1
c	45	145.8	13.3	447	45	AI376292	AI376292 ta61c07.x

ALIGNMENTS

```
RESULT   1
AI104547/c
LOCUS       AI104547      602 bp    mRNA            EST       31-JAN-1999
DEFINITION  EST213836 Normalized rat heart, Bento Soares Rattus sp. cDNA clone
            RHECE94 3' end, mRNA sequence.
ACCESSION   AI104547
VERSION     AI104547.1  GI:3708888
KEYWORDS    EST.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 602)
  AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
  TITLE     Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     Other_ESTs: TC54601
            Contact: Lee, NH
            ATCC
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13-21.
FEATURES             Location/Qualifiers
     source          1. .602
                     /organism="Rattus sp."
                     /db_xref="ATCC (inhost):2025533"
                     /db_xref="taxon:10118"
                     /clone="RHECE94"
                     /clone_lib="Normalized rat heart, Bento Soares"
                     /note="Organ: heart; Vector: pT7T3Pac; Site_1: EcoRI;
                     Site_2: NotI"
BASE COUNT      144 a    171 c    145 g    142 t
ORIGIN


  Query Match           54.1%;  Score 595.6;  DB 43;  Length 602;
  Best Local Similarity  99.3%;  Pred. No. 3.3e-157;
  Matches  598;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy     482 cattcctcaggatctaattggagatttggccagtttggcatttgggagtcaacgccctct 541
           ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     602 CATTCCTCAGGATCTAATTGGAGATTTGGCCAGTTTGGCATTTGGAAGTCAACGCCCTCT 543

Qy     542 tctcgactctgtagcccaacagcagggatcctcgctgcctcacgtgtcttacttccggtg 601
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     542 TCTCGACTCTGTAGCCCAACAGCAGGGATCCTCGCTGCCTCACGTGTCTTACTTCCGGTG 483

Qy     602 gcgggtggacgtggccatctcaaccagcgctcagtcccgctccctgcaaccgagtgttct 661
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     482 CCGGGTGGACGTGGCCATCTCAACCAGCGCTCAGTCCCGCTCCCTGCAACCGAGTGTTCT 423

Qy     662 catgcagctgaagctcacagatggatctgcacaccgcttcgaggtgcccatagccaaatt 721
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     422 CATGCAGCTGAAGCTCACAGATGGATCTGCACACCGCTTCGAGGTGCCCATAGCCAAATT 363

Qy     722 tcaggagctgcggtacagtgtagccttggtccttaaggagatggcagaactggagaagaa 781
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     362 TCAGGAGCTGCGGTACAGTGTAGCCTTGGTCCTTAAGGAGATGGCAGAACTGGAGAAGAA 303

Qy     782 gtgtgagcgcaaactgcaggactgactgaaccctggtactgtgggtgctgaagctggtac 841
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     302 GTGTGAGCGCAAACTGCAGGACTGACTGAACCCTGGTACTGTGGGTGCTGAAGCTGGTAC 243

Qy     842 cagaacacagccccccactggtgatgagcccaactccattgaggtcctgcatgtgagaac 901
           |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db     242 CAGAACACAGCCCCCCACTGGTGATGAGGCCAACTCCATTGAGGTCCTGCATGTGAGAAC 183

Qy     902 gtattttaagtgaaaagacagcgggactttcaggttttgttttatgagtcaacagctggg 961
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     182 GTATTTTAAGTGAAAAGACAGCGGGACTTTCAGGTTTTGTTTTATGAGTCAACAGCTGGG 123

Qy     962 cagggtggcacagtttataatctcagcccttggaagtctgaggctggagaatgggaagtg 1021
           ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     122 CAGGGTGGCACAGTTTATAATCTCAGCCCTTGGAAGGCTGAGGCTGGAGAATGGGAAGTG 63

Qy    1022 taagctgggcctggctttcaagtgaggctcagtgtcgaattaaagaggtaaagcaacta 1081
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62 TAAGCTGGGCCTGGCTTTCATAGTGAGGCTCAGTGTCGAATTAAAGAGGTAAAGCAACTA 3

Qy    1082 tt 1083
           ||
Db       2 TT 1


RESULT   2
AI232721/c
LOCUS       AI232721      680 bp    mRNA            EST       08-JAN-1999
DEFINITION  EST229409 Normalized rat kidney, Bento Soares Rattus sp. cDNA clone
            RKICF61 3' end, mRNA sequence.
```

```
ACCESSION   AI232721
VERSION     AI232721.1  GI:4135792
KEYWORDS    EST.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 680)
  AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
  TITLE     Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     On Oct 30, 1998 this sequence version replaced gi:3816601.
            Other_ESTs: TC60009
            Contact: Lee, NH
            ATCC
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13-21.
FEATURES             Location/Qualifiers
     source          1. .680
                     /organism="Rattus sp."
                     /db_xref="taxon:10118"
                     /clone="RKICF61"
                     /clone_lib="Normalized rat kidney, Bento Soares"
                     /note="Organ: kidney; Vector: pT7T3Pac; Site_1: EcoRI;
                     Site_2: NotI"
BASE COUNT      163 a    187 c    161 g    169 t
ORIGIN
```

```
  Query Match             51.9%;  Score 570.4;  DB 45;  Length 680;
  Best Local Similarity   99.0%;  Pred. No. 4.3e-150;
  Matches  574;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy    511 ccagtttggcatttgggagtcaacgccctcttctcgactctgtagcccaacagcagggat 570
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    680 CCAGTTTGGCATTTGGGAGTCAACGCCCTCTTCTCGACTCTGTAGCCCAACAGCAGGGAA 621

Qy    571 cctcgctgcctcacgtgtcttacttccggtggcgggtggacgtggccatctcaaccagcg 630
          |||||||||||||||||||||||||||||| |||||| ||||||||||||||||||||||
Db    620 CCTCGCTGCCTCACGTGTCTTACTTCCGGTGACGGGTGAACGTGGCCATCTCAACCAGCG 561

Qy    631 ctcagtcccgctccctgcaaccgagtgttctcatgcagctgaagctcacagatggatctg 690
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    560 CTCAGTCCCGCTCCCTGCAACCGAGTGTTCTCATGCAGCTGAAGCTCACAGATGGATCTG 501

Qy    691 cacaccgcttcgaggtgcccatagccaaatttcaggagctgcggtacagtgtagccttgg 750
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    500 CACACCGCTTCGAGGTGCCCATAGCCAAATTTCAGGAGCTGCGGTACAGTGTAGCCTTGG 441

Qy    751 tccttaaggagatggcagaactggagaagaagtgtgagcgcaaactgcaggactgactga 810
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    440 TCCTTAAGGAGATGGCAGAACTGGAGAAGAAGTGTGAGCGCAAACTGCAGGACTGACTGA 381

Qy    811 accctggtactgtgggtgctgaagctggtaccagaacacagccccccactggtgatgagc 870
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    380 ACCCTGGTACTGTGGGTGCTGAAGCTGGTACCAGAACACAGCCCCCCACTGGTGATGAGG 321

Qy    871 ccaactccattgaggtcctgcatgtgagaacgtattttaagtgaaaagacagcgggactt 930
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    320 CCAACTCCATTGAGGTCCTGCATGTGAGAACGTATTTTAAGTGAAAAGACAGCGGGACTT 261

Qy    931 tcaggttttgttttatgagtcaacagctgggcagggtggcacagtttataatctcagccc 990
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    260 TCAGGTTTTGTTTTATGAGTCAACAGCTGGGCAGGGTGGCACAGTTTATAATCTCAGCCC 201

Qy    991 ttggaagtctgaggctggagaatgggaagtgtaagctgggcctggctttcatagtgaggc 1050
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    200 TTGGAAGGCTGAGGCTGGAGAATGGGAAGTGTAAGCTGGGCCTGGCTTTCATAGTGAGGC 141

Qy   1051 tcagtgtcgaattaaagaggtaaagcaactattaaaaaaa 1090
          |||||||||||||||||||||||||||||||||||||||
Db    140 TCAGTGTCGAATTAAAGAGGTAAAGCAACTATTAAAAACA 101
```

```
RESULT   3
AI233681/c
LOCUS       AI233681      462 bp    mRNA            EST       08-JAN-1999
DEFINITION  EST230369 Normalized rat kidney, Bento Soares Rattus sp. cDNA clone
            RKIDI81 3' end, mRNA sequence.
ACCESSION   AI233681
VERSION     AI233681.1  GI:3817561
KEYWORDS    EST.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 462)
  AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
  TITLE     Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     On Jan 17, 1998 this sequence version replaced gi:2044369.
            Contact: Lee, NH
            ATCC
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13-21.
FEATURES             Location/Qualifiers
     source          1. .462
                     /organism="Rattus sp."
                     /db_xref="taxon:10118"
                     /clone="RKIDI81"
                     /clone_lib="Normalized rat kidney, Bento Soares"
                     /note="Organ: kidney; Vector: pT7T3Pac; Site_1: EcoRI;
                     Site_2: NotI"
BASE COUNT      108 a    133 c    100 g    121 t
ORIGIN
```

```
  Query Match             40.6%;  Score 446.6;  DB 43;  Length 462;
  Best Local Similarity   99.1%;  Pred. No. 2.9e-115;
  Matches  449;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy    638 ccgctccctgcaaccgagtgttctcatgcagctgaagctcacagatggatctgcacaccg 697
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    462 CCGCTCCCTGCAACCGAGTGTTCTCATGCAGCTGAAGCTCACAGATGGATCTGCACACCG 403

Qy    698 cttcgaggtgcccatagccaaatttcaggagctgcggtacagtgtagccttggtccttaa 757
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    402 CTTCGAGGTGCCCATAGCCAAACTTCAGGAGCTGCGGTACAGTGTAGCCTTGGTCCTTAA 343

Qy    758 ggagatggcagaactggagaagaagtgtgagcgcaaactgcaggactgactgaaccctgg 817
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    342 GGAGATGGCAGAACTGGAGAAGAAGTGTGAGCGCAAACTGCAGGACTGACTGAACCCTGG 283

Qy    818 tactgtgggtgctgaagctggtaccagaacacagcccccccactggtgatgagcccaactc 877
          |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db    282 TACTGTGGGTGCTGAAGCTGGTACCAGAACACAGCCCCCCACTGGTGATGAGGCCAACTC 223

Qy    878 cattgaggtcctgcatgtgagaacgtattttaagtgaaaagacagcgggactttcaggtt 937
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    222 CATTGAGGTCCTGCATGTGAGAACGTATTTTAAGTGAAAAGACAGCGGGACTTTCAGGTT 163

Qy    938 ttgttttatgagtcaacagctgggcagggtggcacagtttataatctcagcccttggaag 997
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    162 TTGTTTTATGAGTCAACAGCTGGGCAGGGTGGCACAGTTTATAATCTCAGCCCTTGGAAG 103
```

```
Qy    998 tctgaggctggagaatgggaagtgtaagctgggcctggctttcatagtgaggctcagtgt 1057
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    102 GCTGAGGCTGGAGAATGGGAAGTGTAAGCTGGGCCTGGCTTTCATAGTGAGGCTCAGTGT 43

Qy   1058 cgaattaaagaggtaaagcaactattaaaaaaa 1090
          |||||||||||||||||||||||||||||| |
Db     42 CGAATTAAAGAGGTAAAGCAACTATTAAAAACA 10


RESULT   4
AA034568
LOCUS       AA034568      456 bp    mRNA            EST       23-AUG-1996
DEFINITION  mi49a11.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone IMAGE:466844 5', mRNA sequence.
ACCESSION   AA034568
VERSION     AA034568.1  GI:1506431
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 456)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Nov 29, 1993 this sequence version replaced gi:634431.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:280660
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 421.
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:466844"
                     /clone_lib="Soares mouse embryo NbME13.5 14.5"
                     /sex="unknown"
                     /tissue_type="embryo"
                     /dev_stage="13.5-14.5dpc total fetus"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTTTT
                     T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
                     14.5dpc embryos [total RNA provided by  Minoru Ko, Wayne
                     State Univ., from 2 ]; double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector. Library went through one round of
                     normalization, and was constructed by Bento Soares and
                     M.Fatima Bonaldo.  "
BASE COUNT       97 a    134 c    131 g     94 t
ORIGIN


  Query Match           34.9%;  Score 383.4;  DB 27;  Length 456;
  Best Local Similarity  91.8%;  Pred. No. 1.8e-97;
  Matches  405;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;
```

```
Qy     77 atgcaactgtggttaggaccttctcttcggactggtcaagaaacgggaagaaaggatgtc 136
          | || |||||||||||| ||||||||| ||| |||||||||||| || ||| |||||||||
Db     16 AGGCTACTGTGGTTAGTACCTTCTCTCCGGCCTGGTCAAGAAAGGGAAAGGAAGGATGTC 75

Qy    137 tgctttgggggctgcagctccatacttgcaccatcccgctgacagtcacagtggccgggt 196
          ||| |||||||||||| |||||||||| ||||||||  |||||||||||||| ||||||||
Db     76 TGCGTTGGGGGCTGCGGCTCCATACTTACACCATCCAACTGACAGTCACAGCGGCCGGGT 135

Qy    197 cagtttcctgggttcccagccctctccagaagtgacggccgtggctcagctcttgaagga 256
          |||||||||||||||||||||||| |||||||||||| ||||| ||||||||||||||||
Db    136 CAGTTTCCTGGGTTCCCAGCCCTCTGCAGAAGTGACGGCAGTGGCCCAGCTCTTGAAGGA 195

Qy    257 cttagacaggagcaccttcagaaagttgttgaaacttgtagtcggggccctgcatgggaa 316
          |||||||||||||||||||||||||||||||||||| |||||||||||||| ||||||||
Db    196 CTTAGACAGGAGCACCTTCAGAAAGTTGTTGAAACTCGTAGTCGGGGCCCTCCATGGGAA 255

Qy    317 agactgcagagaagctgtggagcaacttggtgccagcgccaacctgtcagaagagcgtct 376
          |||||||||||| || ||| |||| |||||||| ||||||||||||||||| ||||||||
Db    256 AGACTGCAGAGAGGCCGTGCAGCACCTTGGTGCTAGCGCCAACCTGTCAGAGGAGCGTCT 315

Qy    377 ggccgtcctgctggcgggcacacacaccctgctccagcaggctctccggctgccccctgc 436
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    316 GGACGTCCTGCTGGCGGGCACACACACCCTGCTCCAGCAGGCTCTCCGGCTGCCCCCTGC 375

Qy    437 tagtctaaagccagatgccttccaggaagagctccaggaacttggcattcctcaggatct 496
          ||||| || |||||||||||||||| || ||||||||||||||||||| || ||| |
Db    376 CAGTCTGAAACCAGATGCCTTCCAGGACGAACTCCAGGAACTTGGCATTCCCCAAGATAT 435

Qy    497 aattggagatttggccagttt 517
          ||||||||| ||||||||||
Db    436 GATTGGAGATCTGGCCAGTTT 456
```

```
RESULT   5
AI136965/c
LOCUS       AI136965      386 bp    mRNA            EST       05-JUL-1999
DEFINITION  UI-R-C2p-nz-d-10-0-UI.s1 UI-R-C2p Rattus norvegicus cDNA clone
            UI-R-C2p-nz-d-10-0-UI 3', mRNA sequence.
ACCESSION   AI136965
VERSION     AI136965.1  GI:3637742
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 386)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2150747.
            Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence tag present in the cDNA between the NotI site and the
            oligo-dT track served to identify it as a clone from the normalized
            adult Spleen library. cDNA Library Preparation: M. Fatima Bonaldo,
            Ph.D. Clone distribution: clones will be available through Research
            Genetics This clone is also available through the I.M.A.G.E.
            Consortium at LLNL (info@image.llnl.gov). IMAGE ID=1787109 The
            following repetitive elements were found in this cDNA sequence:
            33-145, >PB1D7#SINE/Alu
            Seq primer: M13 Forward
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .386
                     /organism="Rattus norvegicus"
```

/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-C2p-nz-d-10-0-UI"
/clone_lib="UI-R-C2p"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C2p library is a subtracted library derived from the UI-R-C1 library, which is a subtracted library derived from the UI-R-C0 library. The UI-R-C0 library consisted of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, muscle, 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-C2p) was constructed as follows: PCR amplified cDNA inserts from UI-R-C1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the UI-R-C1 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-C2p library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)"
BASE COUNT 87 a 113 c 71 g 115 t
ORIGIN

Query Match 34.8%; Score 382.8; DB 42; Length 386;
Best Local Similarity 99.5%; Pred. No. 2.5e-97;
Matches 384; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy    715 ccaaatttcaggagctgcggtacagtgtagccttggtccttaaggagatggcagaactgg 774
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    386 CCAAATTTCAGGAGCTGCGGTACAGTGTAGCCTTGGTCCTTAAGGAGATGGCAGAACTGG 327

Qy    775 agaagaagtgtgagcgcaaactgcaggactgactgaaccctggtactgtgggtgctgaag 834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    326 AGAAGAAGTGTGAGCGCAAACTGCAGGACTGACTGAACCCTGGTACTGTGGGTGCTGAAG 267

Qy    835 ctggtaccagaacacagccccccactggtgatgagcccaactccattgaggtcctgcatg 894
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    266 CTGGTACCAGAACACAGCCCCCCACTGGTGATGAGGCCAACTCCATTGAGGTCCTGCATG 207

Qy    895 tgagaacgtattttaagtgaaaagacagcgggactttcaggttttgttttatgagtcaac 954
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    206 TGAGAACGTATTTTAAGTGAAAAGACAGCGGGACTTTCAGGTTTTGTTTTATGAGTCAAC 147

Qy    955 agctgggcagggtggcacagtttataatctcagcccttggaagtctgaggctggagaatg 1014
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db    146 AGCTGGGCAGGGTGGCACAGTTTATAATCTCAGCCCTTGGAAGGCTGAGGCTGGAGAATG 87

Qy   1015 ggaagtgtaagctgggcctggctttcatagtgaggctcagtgtcgaattaaagaggtaaa 1074
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     86 GGAAGTGTAAGCTGGGCCTGGCTTTCATAGTGAGGCTCAGTGTCGAATTAAAGAGGTAAA 27

Qy   1075 gcaactattaaaaaaaaaaaaaaaa 1100
          ||||||||||||||||||||||||
Db     26 GCAACTATTAAAAAAAAAAAAAAAA 1
```

RESULT 6
AI048873
LOCUS AI048873 482 bp mRNA EST 08-JUL-1998
DEFINITION uc76f02.y1 Sugano mouse liver mlia Mus musculus cDNA clone IMAGE:1431579 5', mRNA sequence.
ACCESSION AI048873
VERSION AI048873.1 GI:3297160
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 482)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Sekhon,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Jan 14, 1998 this sequence version replaced gi:1798596.
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:915647
Seq primer: custom primer used
High quality sequence stop: 463.
FEATURES Location/Qualifiers
source 1. .482
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:1431579"
/clone_lib="Sugano mouse liver mlia"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTTGGCCTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end primer CGACCTGCAGCTCGAGCACA."
BASE COUNT 102 a 144 c 136 g 100 t
ORIGIN

Query Match 34.2%; Score 376.6; DB 41; Length 482;
Best Local Similarity 91.1%; Pred. No. 1.5e-95;
Matches 400; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

```
Qy     71 tttgtaatgcaactgtggttaggaccttctcttcggactggtcaagaaacgggaagaaag 130
          ||  |  ||| |||||||||||| ||||||||| ||| ||| |||||||| || ||| |||
Db     44 TTCCTTTTGCTACTGTGGTTAGTACCTTCTCTCCGGGCTGATCAAGAAAGGGAAAGGAAG 103

Qy    131 gatgtctgctttgggggctgcagctccatacttgcaccatcccgctgacagtcacagtgg 190
          |||||||| |||||||||| |||||||||| |||||||  ||||||||||||| ||
Db    104 GATGTCTGCGTTGGGGGCTGCGGCTCCATACTTACACCATCCAACTGACAGTCACAGCGG 163

Qy    191 ccgggtcagtttcctgggttcccagccctctccagaagtgacggccgtggctcagctctt 250
          |||||||||||||||||||||||||||||||| |||||||||||| ||||| ||||||||
Db    164 CCGGGTCAGTTTCCTGGGTTCCCAGCCCTCTGCAGAAGTGACGGCAGTGGCCCAGCTCTT 223

Qy    251 gaaggacttagacaggagcaccttcagaaagttgttgaaacttgtagtcggggccctgca 310
          |||||||||||||||||||||||||||||||||||||||| |||||||||||||| ||
Db    224 GAAGGACTTAGACAGGAGCACCTTCAGAAAGTTGTTGAAACTCGTAGTCGGGGCCCTCCA 283
```

```
Qy    311 tgggaaagactgcagagaagctgtggagcaacttggtgccagcgccaacctgtcagaaga 370
          ||||||||||||||||||| || ||| |||| ||||||||||||||||||||||||||| ||
Db    284 TGGGAAAGACTGCAGAGAGGCCGTGCAGCACCTTGGTGCCAGCGCCAACCTGTCAGAGGA 343

Qy    371 gcgtctggccgtcctgctggcgggcacacacaccctgctccagcaggctctccggctgcc 430
          ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    344 GCGTCTGGCCGTCCTGCTGGCGGGCACACACACCCTGCTCCAGCAGGCTCTTCGGCTGCC 403

Qy    431 ccctgctagtctaaagccagatgccttccaggaagagctccaggaacttggcattcctca 490
          |||||| ||||| || |||||||||||||||| || |||||||||||||||||||| ||
Db    404 CCCTGCCAGTCTGAAACCAGATGCCTTCCAGGACGAACTCCAGGAACTTGGCATTCCCCA 463

Qy    491 ggatctaattggagatttg 509
          ||| | |||||||||| ||
Db    464 AGATATGATTGGAGATCTG 482
```

RESULT 7
AA032667
LOCUS AA032667 444 bp mRNA EST 22-AUG-1996
DEFINITION mi32b07.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA clone IMAGE:465205 5', mRNA sequence.
ACCESSION AA032667
VERSION AA032667.1 GI:1505234
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 444)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Nov 29, 1993 this sequence version replaced gi:636534.
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:279021
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 434.
FEATURES Location/Qualifiers
source 1. .444
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:465205"
/clone_lib="Soares mouse embryo NbME13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTTTT T 3'], on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo. "

BASE COUNT 95 a 133 c 128 g 88 t
ORIGIN

Query Match 33.9%; Score 373; DB 27; Length 444;
Best Local Similarity 91.8%; Pred. No. 1.5e-94;
Matches 394; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

```
Qy     77 atgcaactgtggttaggaccttctcttcggactggtcaagaaacgggaagaaaggatgtc 136
          | || |||||||||||| ||||||||| ||| |||||||||||| || ||| |||||||||
Db     16 AGGCTACTGTGGTTAGTACCTTCTCTCCGGCCTGGTCAAGAAAGGGAAAGGAAGGATGTC 75

Qy    137 tgctttgggggctgcagctccatacttgcaccatcccgctgacagtcacagtggccgggt 196
          ||| |||||||||||| |||||||||| |||||||| |||||||||||||| ||||||||
Db     76 TGCGTTGGGGGCTGCGGCTCCATACTTACACCATCCAACTGACAGTCACAGCGGCCGGGT 135

Qy    197 cagtttcctgggttcccagccctctccagaagtgacggccgtggctcagctcttgaagga 256
          |||||||||||||||||||||||| |||||||||||| ||||| ||||||||||||||||
Db    136 CAGTTTCCTGGGTTCCCAGCCCTCTGCAGAAGTGACGGCAGTGGCCCAGCTCTTGAAGGA 195

Qy    257 cttagacaggagcaccttcagaaagttgttgaaacttgtagtcggggccctgcatgggaa 316
          |||||||||||||||||||||||||||||||| |||||||||||||| ||||||||||||
Db    196 CTTAGACAGGAGCACCTTCAGAAAGTTGTTGAAACTCGTAGTCGGGGCCCTCCATGGGAA 255

Qy    317 agactgcagagaagctgtggagcaacttggtgccagcgccaacctgtcagaagagcgtct 376
          |||||||||||| || ||| |||| |||||||  |||||||||||||||||| ||||||||
Db    256 AGACTGCAGAGAGGCCGTGCAGCACCTTGGTCGCAGCGCCAACCTGTCAGAGGAGCGTCT 315

Qy    377 ggccgtcctgctggcgggcacacacaccctgctccagcaggctctccggctgccccctgc 436
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    316 GGCCGTCCTGCTGGCGGGCACACACACCCTGCTCCAGCAGGCTCTCCGGCTGCCCCCTGC 375

Qy    437 tagtctaaagccagatgccttccaggaagagctccaggaacttggcattcctcaggatct 496
          ||||| || |||||||||||||||| || |||||||||||||||||||| || ||| |
Db    376 CAGTCTGAAACCAGATGCCTTCCAGGACGAACTCCAGGAACTTGGCATTCCCCAAGATAT 435

Qy    497 aattggaga 505
          |||||||||
Db    436 GATTGGAGA 444
```

RESULT 8
AI154070
LOCUS AI154070 454 bp mRNA EST 30-SEP-1998
DEFINITION ud76a06.r1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone IMAGE:1476754 5', mRNA sequence.
ACCESSION AI154070
VERSION AI154070.1 GI:3682539
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 454)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Jan 14, 1998 this sequence version replaced gi:1797639.
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:925110
Seq primer: -28m13 rev2 ET from Amersham

```
          High quality sequence stop: 452.
FEATURES             Location/Qualifiers
     source          1. .454
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1476754"
                     /clone_lib="Soares_mammary_gland_NMLMG"
                     /sex="female (lactating)"
                     /tissue_type="mammary gland"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from mammary
                     gland tissue from a lactating female, and was then primed
                     with a Not I - oligo(dT) primer. Double-stranded  cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I  and cloned into the Not I and Eco RI sites of the
                     modified pT7T3 vector.  Library is normalized.  Library
                     was constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       96 a    131 c    134 g     93 t
ORIGIN


  Query Match             32.9%;  Score 362;  DB 42;  Length 454;
  Best Local Similarity   91.6%;  Pred. No. 1.9e-91;
  Matches  416;  Conservative    0;  Mismatches   35;  Indels    3;  Gaps    3;

Qy      77 atgcaactgtggttaggaccttctcttcggactggtcaagaaacgggaagaaaggatgtc 136
           | || |||||||||||  |||||||||  |||  ||||||||||||  ||  |||  |||||||||
Db       4 AGGCTACTGTGGTTAGTACCTTCTCTCCGGCCTGGTCAAGAAAGGGAAAGGAAGGATGTC  63

Qy     137 tgctttgggggctgcagctccatacttgcaccatcccgctgacagtcacagtggccgggt 196
           |||  ||||||||||| ||||||||||  ||||||||  ||||||||||||| ||||||||
Db      64 TGCGTTGGGGGCTGCGGCTCCATACTTACACCATCCAACTGACAGTCACAGCGGCCGGGT 123

Qy     197 cagtttcctgggttcccagccctctccagaagtgacggccgtggctcagctcttgaagga 256
           ||||||||||||||||||||||||  |||||||||||||  |||||  ||||||||||||||
Db     124 CAGTTTCCTGGGTTCCCAGCCCTCTGCAGAAGTGACGGCAGTGGCCCAGCTCTTGAAGGA 183

Qy     257 cttagacaggagcaccttcagaaagttgttgaaacttgtagtcggggccctgcatgggaa 316
           |||||||||||||||||||||||||||||||||||||||  ||||||||||||||  ||||||||
Db     184 CTTAGACAGGAGCACCTTCAGAAAGTTGTTGAAACTCGTAGTCGGGGCCCTCCATGGGAA 243

Qy     317 agactgcagagaagctgtggagcaacttggtgccagcgccaacctgtcagaagagcgtct 376
           ||||||||||||  ||  |||  ||||  |||||||  ||||||||||||||||||  ||||||||
Db     244 AGACTGCAGAGAGGCCGTGCAGCACCTTGGTG-CAGCGCCAACCTGTCAGAGGAGCGTCT 302

Qy     377 ggccgtcctgctggcgggcacacacaccctgctccagcaggctctccggctgccccctgc 436
           ||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||||||||
Db     303 GGCCGTCCTGCTGGCGGGCACACACACCCTGCTCCAGCA-GCTCTCCGGCTGCCCCCTGC 361

Qy     437 tagtctaaagccagatgccttccaggaagagctccaggaacttggcattcctcaggatct 496
           |||||  ||  ||||||||||||||||  ||  ||||||||||||||||||||  ||  |||  |
Db     362 CAGTCTGAAACCAGATGCCTTCCAGGACGAACTCCAGGAACTTGGCATTCC-CAAGATAT 420

Qy     497 aattggagatttggccagtttggcatttgggagt 530
           |||||||||  ||||||||||  ||||||||||
Db     421 GATTGGAGATCTGGCCAGTTTGCGATTTGGGAGT 454


RESULT   9
AW434638/c
LOCUS       AW434638      345 bp    mRNA            EST       11-FEB-2000
DEFINITION  UI-R-BJ0p-afi-e-01-0-UI.s1 UI-R-BJ0p Rattus norvegicus cDNA clone
            UI-R-BJ0p-afi-e-01-0-UI 3', mRNA sequence.
ACCESSION   AW434638
VERSION     AW434638.1  GI:6965945
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 345)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     On May 18, 1998 this sequence version replaced gi:3137829.
            Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            normalized atrium at 15 dpc library cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: clones will be available through
            Research Genetics (www.resgen.com) The following repetitive
            elements were found in this cDNA sequence: 34-146, >PB1D7#SINE/Alu
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .345
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-BJ0p-afi-e-01-0-UI"
                     /clone_lib="UI-R-BJ0p"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BJ0p
                     library is a subtracted library derived from the UI-R-AA1,
                      UI-R-AB1, UI-R-AC1, UI-R-AD1, UI-R-AE1, UI-R-AF1, and
                     UI-R-AG1 libraries. These libraries represent tissues from
                     rat atrium at 16.5 dpc, ventricle  at 16.5 dpc, AV canal
                     at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc, AV
                     canal at 15 dpc, and ventricle at 13 dpc.  The tag is a
                     string of 5-6 nucleotides present between the Not I site
                     and the oligo-dT track. The library was constructed as
                     described by Bonaldo, Lennon and Soares,  Genome Research
                     6: 791-806, 1996.
                     TAG_LIB=UI-R-BJ0p
                     TAG_TISSUE=atrium at 15 dpc
                     TAG_SEQ=GAAGC"
BASE COUNT       77 a    101 c     63 g    104 t
ORIGIN


  Query Match             30.4%;  Score 334.4;  DB 81;  Length 345;
  Best Local Similarity   98.3%;  Pred. No. 1e-83;
  Matches  338;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy     757 aggagatggcagaactggagaagaagtgtgagcgcaaactgcaggactgactgaaccctg 816
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     345 AGGAGATGGCAGAACTGGAGAAGAAGTGTGAGCGCAAACTCCAGGACTGACTGAACCCTG 286

Qy     817 gtactgtgggtgctgaagctggtaccagaacacagccccccactggtgatgagcccaact 876
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db     285 GTACTGTGGGTGCTGAAGCTGGTACCAGAACACAGCCCCCCACTGGTGATGAGGCCAACT 226

Qy     877 ccattgaggtcctgcatgtgagaacgtattttaagtgaaaagacagcgggactttcaggt 936
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     225 CCATTGAGGTCCTGCATGTGAGAACGTATTTTAAGTGAAAAGACAGCGGGACTTTCAGGT 166

Qy     937 tttgttttatgagtcaacagctgggcagggtggcacagtttataatctcagcccttggaa 996
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     165 TTTGTTTTATGAGTCAACAGCTGGGCAGGGTGGCACAGTTTATAATCTCAGCCCTTGGAA 106

Qy     997 gtctgaggctggagaatgggaagtgtaagctgggcctggctttcatagtgaggctcagtg 1056
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    105 GGCTGAGGCTGGAGAATGGGAAGTGTAAGCTGGGCCTGGCTTTCATAGTGAGGCTCAGTG 46

Qy   1057 tcgaattaaagaggtaaagcaactattaaaaaaaaaaaaaaaaa 1100
          ||||||||||||||||||||| || || ||||||||||||||||
Db     45 TCGAATTAAAGAGGTAAAGCACCTTTTTAAAAAAAAAAAAAAAA 2


RESULT  10
AQ055548/c
LOCUS       AQ055548      540 bp    DNA             GSS       30-JUL-1998
DEFINITION  CIT-HSP-2338P5.TR CIT-HSP Homo sapiens genomic clone 2338P5,
            genomic survey sequence.
ACCESSION   AQ055548
VERSION     AQ055548.1  GI:3352154
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 540)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
            Simon,M. and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building (1998)
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: CIT-HSP-2338P5.TF
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .540
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="2338P5"
                     /clone_lib="CIT-HSP"
                     /sex="Male"
                     /cell_type="Sperm"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
BASE COUNT      107 a    170 c    166 g     97 t
ORIGIN


  Query Match             29.8%;  Score 327.4;  DB 94;  Length 540;
  Best Local Similarity   77.4%;  Pred. No. 1.2e-81;
  Matches  411;  Conservative    0;  Mismatches  116;  Indels    4;  Gaps    1;

Qy    150 gcagctccatacttgcaccatcccgctgacagtcacagtggccgggtcagtttcctgggt 209
          ||| |||||||| |||| ||||| | | | ||||||||||||||| || |||||| ||||
Db    540 GCAACTCCATACATGCATCATCCTGGTAATAGTCACAGTGGCCGAGTGAGTTTCTTGGGG 481

Qy    210 tcccag----ccctctccagaagtgacggccgtggctcagctcttgaaggacttagacag 265
          |||||    ||| ||||||| ||| | ||  |||| | |||  |   |||| ||||||| 
Db    480 GCCCAGCTTTCCCACTCCAGAGGTGGCAGCAATGGCCCGGCTACTAGGGGACCTAGACAG 421

Qy    266 gagcaccttcagaaagttgttgaaacttgtagtcggggccctgcatgggaaagactgcag 325
          |||||| |||||||||||| |||| |||| ||| |   |||||| ||||| |||||| |
Db    420 GAGCACGTTCAGAAAGTTGCTGAAGTTTGTGGTCAGCAGCCTGCAGGGGAAGGACTGCCG 361

Qy    326 agaagctgtggagcaacttggtgccagcgccaacctgtcagaagagcgtctggccgtcct 385
          ||| |||||| |||  ||||| | |||||||||||||| | || ||||  ||||  | |||
Db    360 AGAGGCTGTGCAGCGTCTTGGGGTCAGCGCCAACCTGCCGGAGGAGCAGCTGGGTGCCCT 301

Qy    386 gctggcgggcacacacaccctgctccagcaggctctccggctgccccctgctagtctaaa 445
          |||||| |||| ||||| |||||||||||||| ||||| ||||||||  | || || ||
Db    300 GCTGGCAGGCATGCACACACTGCTCCAGCAGGCCCTCCGTCTGCCCCCCACCAGCCTGAA 241

Qy    446 gccagatgccttccaggaagagctccaggaacttggcattcctcaggatctaattggaga 505
          ||| ||  ||||| |||  |||||||||| ||  |||| || || || ||  | || ||
Db    240 GCCTGACACCTTCAGGGACCAGCTCCAGGAGCTCTGCATCCCCCAAGACCTGGTCGGGGA 181

Qy    506 tttggccagtttggcatttgggagtcaacgccctcttctcgactctgtagcccaacagca 565
          |||||||  ||| |||||||||| || || || || || || ||||| |||||  |||||
Db    180 CTTGGCCAGCGTGGTATTTGGGAGCCAGCGGCCCCTCCTTGATTCTGTGGCCCAGCAGCA 121

Qy    566 gggatcctcgctgcctcacgtgtcttacttccggtggcgggtggacgtggccatctcaac 625
          |||  ||| |||||| || ||  || |||| ||||||||||||||| || || |||||  ||
Db    120 GGGGGCCTGGCTGCCGCATGTTGCTGACTTTCGGTGGCGGGTGGATGTAGCAATCTCCAC 61

Qy    626 cagcgctcagtcccgctccctgcaaccgagtgttctcatgcagctgaagct 676
          |||.|| | | | |||||||||||| ||||| || || ||||||||||||||
Db     60 CAGTGCCCTGGCTCGCTCCCTGCAGCCGAGCGTCCTGATGCAGCTGAAGCT 10


RESULT  11
AA065442
LOCUS       AA065442      394 bp    mRNA            EST       03-FEB-1997
DEFINITION  ml38h02.r1 Stratagene mouse testis (#937308) Mus musculus cDNA
            clone IMAGE:514323 5', mRNA sequence.
ACCESSION   AA065442
VERSION     AA065442.1  GI:1562920
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 394)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Jan 25, 1995 this sequence version replaced gi:637728.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:308171
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 392.
FEATURES             Location/Qualifiers
     source          1. .394
                     /organism="Mus musculus"
                     /strain="Inbred CD-1"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:514323"
                     /clone_lib="Stratagene mouse testis (#937308)"
                     /sex="males"
                     /tissue_type="testis"
                     /dev_stage="10-12 week old"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: testis; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. Average insert size: 1.0 kb; Uni-ZAP XR Vector;
                     -5' adaptor sequence: 5' GAATTCGGCACGAG 3'; -3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT       79 a    121 c    117 g     77 t
ORIGIN
```

```
  Query Match             27.9%;  Score 307.4;  DB 28;  Length 394;
  Best Local Similarity   92.6%;  Pred. No. 4.4e-76;
  Matches  323;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy      77 atgcaactgtggttaggaccttctcttcggactggtcaagaaacgggaagaaaggatgtc 136
           | || |||||||||||| |||||||||| ||| |||||||||||| || ||| |||||||||
Db      46 AGGCTACTGTGGTTAGTACCTTCTCTCCGGCCTGGTCAAGAAAGGGAAAGGAAGGATGTC 105

Qy     137 tgctttgggggctgcagctccatacttgcaccatcccgctgacagtcacagtggccgggt 196
           ||| |||||||||||  |||||||||||  ||||||||  ||||||||||||| ||||||||
Db     106 TGCGTTGGGGGCTGCGGCTCCATACTTACACCATCCAACTGACAGTCACAGCGGCCGGGT 165

Qy     197 cagtttcctgggttcccagccctctccagaagtgacggccgtggctcagctcttgaagga 256
           |||||||||||||||||||||||||||||| |||||||||||||  ||||| ||||||||||||||
Db     166 CAGTTTCCTGGGTTCCCAGCCCTCTGCAGAAGTGACGGCAGTGGCCCAGCTCTTGAAGGA 225

Qy     257 cttagacaggagcaccttcagaaagttgttgaaacttgtagtcggggccctgcatgggaa 316
           |||||||||||||||||||||||||||||||||||||||||||||| |||||||||| ||||||||
Db     226 CTTAGACAGGAGCACCTTCAGAAAGTTGTTGAAACTCGTAGTCGGGGCCCTCCATGGGAA 285

Qy     317 agactgcagagaagctgtggagcaacttggtgccagcgccaacctgtcagaagagcgtct 376
           |||||||||||| || ||| |||| ||||||  |||||||||||||||||| ||||||||
Db     286 AGACTGCAGAGAGGCCGTGCAGCACCTTGGTCGCAGCGCCAACCTGTCAGAGGAGCGTCT 345

Qy     377 ggccgtcctgctggcgggcacacacaccctgctccagcaggctctccgg 425
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db     346 GGCCGTCCTGCTGGCGGGCACACACACCCTGCTCCAGCAGGCTCTCCGG 394
```

```
RESULT  12
W83837
LOCUS       W83837          446 bp    mRNA            EST       12-SEP-1996
DEFINITION  mf26c06.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone IMAGE:406186 5', mRNA sequence.
ACCESSION   W83837
VERSION     W83837.1  GI:1540621
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 446)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1394997.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:249954
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 444.
FEATURES             Location/Qualifiers
     source          1. .446
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:406186"
                     /clone_lib="Soares mouse embryo NbME13.5 14.5"
                     /sex="unknown"
                     /tissue_type="embryo"
                     /dev_stage="13.5-14.5dpc total fetus"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTTT
                     T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
                     14.5dpc embryos [total RNA provided by  Minoru Ko, Wayne
                     State Univ., from 2 ]; double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector. Library went through one round of
                     normalization, and was constructed by Bento Soares and
                     M.Fatima Bonaldo. "
BASE COUNT       98 a    108 c    137 g    103 t
ORIGIN
```

```
  Query Match             27.5%;  Score 302.4;  DB 27;  Length 446;
  Best Local Similarity   84.3%;  Pred. No. 1.2e-74;
  Matches  398;  Conservative    0;  Mismatches   46;  Indels   28;  Gaps    4;

Qy     505 atttggccagtttggcatttgggagtcaacgccctcttctcgactctgtagcccaacagc 564
           || |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db       1 ATCTGGCCAGTTTGGCATTTGGGAGTCAACGCCCTCTTCTCGACTCTGTAGCCCAGCAGC 60

Qy     565 agggatcctcgctgcctcacgtgtcttacttccggtggcgggtggacgtggccatctcaa 624
           |||||||||||||||||  |||||  | |||||||||||||||| ||||||||||||||
Db      61 AGGGATCCTCGCTGCCTCGTGTGTCGAATTTCCGGTGGCGGGTGGATGTGGCCATCTCAA 120

Qy     625 ccagcgctcagtcccgctccctgcaaccgagtgttctcatgcagctgaagctcacagatg 684
           |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CCAGTGCTCAGTCCCGCTCCCTGCAACCGAGTGTTCTCATGCAGCTGAAGCTCACAGATG 180

Qy     685 gatctgcacaccgcttcgaggtgcccatagccaaatttcaggagctgcggtacagtgtag 744
           |||| |||||||||||||||||||||||||||||| ||||||||||| ||||| || |
Db     181 GATCGGCACACCGCTTCGAGGTGCCCATAGCCAAATTCCAGGAGCTGCGCTACAGCGTGG 240

Qy     745 ccttggtccttaaggagatggcagaactggagaagaagtgtgagcgcaaactgcaggact 804
           | |||||||||||||||||||||||| |||||||  ||||||||||||||||||||
Db     241 CGTTGGTCCTTAAGGAGATGGCAGAGCTGGAGAGAAAGTGTGAGCGCAAACTGCAG---- 297

Qy     805 gactgaaccctggtactgt-gggtgctgaagctggtaccagaacacagccccccactggt 863
           |||||||||||||||||| |||||||||||||||||||||||| |||  || |||||||
Db     297 GACTGAACCCTGGTACTGTGGGGTGCTGAAGCTGGTACCAGAACCCAG--CCGCACTGGT 354

Qy     864 gatgagcccaactccattgaggtcctgcatgtgagaacgtattttaagtgaaaagacagc 923
           |||||                          | |||| || || ||||||||||||||||||||
Db     355 GATGA--------------------GGATGTAAGGACATATTTTAAGTGAAAAGACAGC 393

Qy     924 gggactttcaggttttgttttatgagtcaacagctgggcagggtggcacagt 975
           ||   ||  || |||  |||   |   |||||||||| || | | | |
Db     394 TGGGACTTTTGGGTTTTGTTTTACAAGTAACAGCTGGGCTAGGCTGGAGAAT 445
```

```
RESULT  13
C77953/c
LOCUS       C77953          569 bp    mRNA            EST       25-JUN-1998
DEFINITION  C77953 Mouse 3.5-dpc blastocyst cDNA Mus musculus cDNA clone
            J0040D09 3', mRNA sequence.
ACCESSION   C77953
VERSION     C77953.1  GI:2518283
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 569)
  AUTHORS   Ko,M.S.H., Kitchen,J.R., Wang,X., Wang,X., Threat,T.A., Sun,T.,
            Grahovac,M.J., Mason,S., Lim,M.K., Paonessa,P.D., Sauls,A.D. and
            Doi,H.
  TITLE     Systematic analyses of genes expressed in 3.5-dpc mouse blastocyst
            (The ERATO/Doi Project at Wayne State University)
  JOURNAL   Unpublished (1997)
```

```
COMMENT     On Apr 18, 1995 this sequence version replaced gi:775302.
            Contact: Hirofumi Doi
            Doi Bioasymmetry Project, ERATO
            Japan Science and Technology Corporation (JST)
            WBG Marive East 12F, 2-6 Nakase, Mihama-ku, Chiba 261-71, Japan
            Email: hd@bioa.jst.go.jp.
FEATURES             Location/Qualifiers
     source          1. .569
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="J0040D09"
                     /clone_lib="Mouse 3.5-dpc blastocyst cDNA"
                     /tissue_type="blastocyst"
                     /dev_stage="3.5-dpc"
BASE COUNT      133 a    169 c    131 g    134 t      2 others
ORIGIN
```

```
  Query Match           27.0%;  Score 296.6;  DB 36;  Length 569;
  Best Local Similarity  80.6%;  Pred. No. 5.6e-73;
  Matches  510;  Conservative    0;  Mismatches   46;  Indels   77;  Gaps   11;

Qy    460 aggaagagctccaggaacttggcattcctc--aggatctaattggagatttggccagttt 517
          |||| || ||||||||||||||||||||||  |  ||  | | |||||||||| ||||||||||
Db    569 AGGACGAACTCCAGGAACTTGGCATTCCCCCAAGATTATGATTGGAGATCTGGCCAGTTT 510

Qy    518 ggcatttgggagtcaacgccc-tcttctcgactctgtagcccaacagcagggat-cctcg 575
          |||||||||| ||||||||| | |||||||||||||||||| |||||||||| |||||
Db    509 GGCATTTGGGANTCAACGCCCTTNTTCTCGACTCTGTAGCCCAGCAGCAGGGATCCCTCG 450

Qy    576 ctgcctcacgtgtcttacttccggtggcgggtggacgtggccatctcaaccagcgctcag 635
          |||||||  |||||  | |||||||||||||||| ||||||||||||||||| ||||||
Db    449 CTGCCTCGTGTGTCGAATTTCCGGTGGCGGGTGGATGTGGCCATCTCAACCAGTGCTCAG 390

Qy    636 tcccgctccctgcaaccgagtgttctcatg-cagctgaagctcacagatggatctgcaca 694
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||| |||||
Db    389 TCCCGCTCCCTGCAACCGAGTGTTCTCATGCCAGCTGAAGCTCACAGATGGATCGGCACA 330

Qy    695 ccgcttcgaggtgcccatagccaaatttcaggagctgcggtacagtgtagccttggtcct 754
          |||||||||||||||||||||||||||||| ||||||||||  ||||| || || ||||||||
Db    329 CCGCTTCGAGGTGCCCATAGCCAAATTCCAGGAGCTGCTCTACAGCGTGGCGTTGGTCCT 270

Qy    755 taaggagatggcagaactggagaagaagtgtgagcgcaaactgcaggactgactgaaccc 814
          |||||||||||||||| |||||||  ||||||||||||||||||||||    ||||||||||
Db    269 TAAGGAGATGGCAGAGCTGGAGAGAAAGTGTGAGCGCAAACTGCAG----GACTGAACCC 214

Qy    815 tggtactgt-gggtgctgaagctggtaccagaacacagccccccactggtgatgagccca 873
          ||||||||| |||||||||||||||||||||||||| |||  ||  ||||||||||||
Db    213 TGGTACTGTGGGGTGCTGAAGCTGGTACCAGAACCCAG--CCGCACTGGTGATGA----- 160

Qy    874 actccattgaggtcctgcatgtgagaacgtattttaagtgaaaagacagc-gggactttc 932
                      | |||| ||||| ||||||||||||||||||||||| |||||||||
Db    160 ----------------GGATGTAAGAACATATTTTAAGTGAAAAGACAGCTGGGACTTTT 117

Qy    933 aggttttgttttatgagtcaacagctgggcagggtggcacagtttataatctcagccctt 992
          ||||||||||||  ||| ||||||||||
Db    116 GGGTTTTGTTTTACAAGTAAACAGCTGGG----------------------------- 87

Qy    993 ggaagtctgaggctggagaatgggaagtgtaagctgggcctggc------tttcatagtg 1046
          |||||||||||||||| ||| ||||||   ||||||||      || | |||||
Db     87 ------CTGAGGCTGGAGAATGTGAAATGTAAGGCCGGCCTGGCTGGGTTTTACGTAGTG 34

Qy   1047 aggctcagtgtcgaattaaagaggtaaagcaac 1079
          ||||||||| |||||||||||||||||||||||
Db     33 AGGCTCAGTTTCGAATTAAAGAGGTAAAGCAAC 1
```

```
RESULT  14
AA450933
LOCUS       AA450933      374 bp    mRNA            EST       04-JUN-1997
DEFINITION  vf85a11.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone
            IMAGE:850556 5', mRNA sequence.
ACCESSION   AA450933
VERSION     AA450933.1  GI:2164603
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 374)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1394992.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:502708
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 348.
FEATURES             Location/Qualifiers
     source          1. .374
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:850556"
                     /clone_lib="Soares_mammary_gland_NbMMG"
                     /sex="male"
                     /tissue_type="mammary gland"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Organ: mammary gland; Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                     oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
BASE COUNT      77 a    113 c    112 g     72 t
ORIGIN
```

```
  Query Match           26.9%;  Score 296.4;  DB 33;  Length 374;
  Best Local Similarity  92.3%;  Pred. No. 5.5e-73;
  Matches  312;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy     77 atgcaactgtggttaggaccttctcttcggactggtcaagaaacgggaagaaaggatgtc 136
          | || |||||||||| ||||||||| ||| |||||||||||| || ||| |||||||||
Db     37 AGGCTACTGTGGTTAGTACCTTCTCTCCGGCCTGGTCAAGAAAGGGAAAGGAAGGATGTC 96

Qy    137 tgctttgggggctgcagctccatacttgcaccatcccgctgacagtcacagtggccgggt 196
          ||| |||||||||| |||||||||| |||||||| |||||||||||| |||||||||
Db     97 TGCGTTGGGGGCTGCGGCTCCATACTTACACCATCCAACTGACAGTCACAGCGGCCGGGT 156

Qy    197 cagtttcctgggttcccagccctctccagaagtgacggccgtggctcagctcttgaagga 256
          |||||||||||||||||||||||| |||||||||||| ||||| ||||||||||||||
Db    157 CAGTTTCCTGGGTTCCCAGCCCTCTGCAGAAGTGACGGCAGTGGCCCAGCTCTTGAAGGA 216

Qy    257 cttagacaggagcaccttcagaaagttgttgaaacttgtagtcggggccctgcatgggaa 316
          |||||||||||||||||||||||||||||||||||||||||||| |||||||||||| ||||||||
Db    217 CTTAGACAGGAGCACCTTCAGAAAGTTGTTGAAACTCGTAGTCGGGGCCCTCCATGGGAA 276
```

```
Qy    317 agactgcagagaagctgtggagcaacttggtgccagcgccaacctgtcagaagagcgtct 376
          ||||||||||||  ||  |||  ||||  ||||||||  ||||||||||||||||  ||||||||
Db    277 AGACTGCAGAGAGGCCGTGCAGCACCTTGGTGCAGGCGCCAACCTGTCAGAGGAGCGTCT 336

Qy    377 ggccgtcctgctggcgggcacacacaccctgctccagc 414
          |||||||||||||||||||||||||||||||||||||||
Db    337 GGCCGTCCTGCTGGCGGGCACACACACCCTGCTCCAGC 374


RESULT  15
C78711/c
LOCUS       C78711         574 bp    mRNA            EST       25-JUN-1998
DEFINITION  C78711 Mouse 3.5-dpc blastocyst cDNA Mus musculus cDNA clone
            J0053G06 3', mRNA sequence.
ACCESSION   C78711
VERSION     C78711.1  GI:2519041
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 574)
  AUTHORS   Ko,M.S.H., Kitchen,J.R., Wang,X., Wang,X., Threat,T.A., Sun,T.,
            Grahovac,M.J., Mason,S., Lim,M.K., Paonessa,P.D., Sauls,A.D. and
            Doi,H.
  TITLE     Systematic analyses of genes expressed in 3.5-dpc mouse blastocyst
            (The ERATO/Doi Project at Wayne State University)
  JOURNAL   Unpublished (1997)
COMMENT     On Dec 3, 1996 this sequence version replaced gi:1119170.
            Contact: Hirofumi Doi
            Doi Bioasymmetry Project, ERATO
            Japan Science and Technology Corporation (JST)
            WBG Marive East 12F, 2-6 Nakase, Mihama-ku, Chiba 261-71, Japan
            Email: hd@bioa.jst.go.jp.
FEATURES             Location/Qualifiers
     source          1. .574
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="J0053G06"
                     /clone_lib="Mouse 3.5-dpc blastocyst cDNA"
                     /tissue_type="blastocyst"
                     /dev_stage="3.5-dpc"
BASE COUNT      133 a    170 c    136 g    133 t      2 others
ORIGIN


  Query Match           26.1%;  Score 286.6;  DB 36;  Length 574;
  Best Local Similarity  80.3%;  Pred. No. 3.7e-70;
  Matches  500;  Conservative    0;  Mismatches   46;  Indels   77;  Gaps   11;

Qy    470 ccaggaacttggcattcctcaggatctaattggagatttggccagtttgg--catttggg 527
          ||||||||||||||||||  ||  |||  |  |||||||||  |||||    ||  |  ||||||||
Db    559 CCAGGAACTTGGCATTCCCCANGATATGATTGGAGATCTGGCCCAGTTTGGCCATTTGGG 500

Qy    528 agtcaacgccctcttctcgactctgtag-cccaacagcagggat-cctcgctgcctcacg 585
          ||||||||||  ||||||||||||||||||  ||||  ||||||||||  ||||||||||  |
Db    499 AGTCAACGCCNTCTTCTCGACTCTGTAGCCCCAGCAGCAGGGATCCCTCGCTGCCTCGTG 440

Qy    586 tgtcttacttccggtggcgggtggacgtggccatctcaaccagcgctcagtcccgctccc 645
          ||||  |  ||||||||||||||||  ||||||||||||||||  ||||||||||||||||
Db    439 TGTCGAATTTCCGGTGGCGGGTGGATGTGGCCATCTCAACCAGTGCTCAGTCCCGCTCCC 380

Qy    646 tgcaaccgagtgttctcatg-cagctgaagctcacagatggatctgcacaccgcttcgag 704
          ||||||||||||||||||  ||||||||||||||||||||||||  ||||||||||||||||
Db    379 TGCAACCGAGTGTTCTCATGCCAGCTGAAGCTCACAGATGGATCGGCACACCGCTTCGAG 320

Qy    705 gtgcccatagccaaatttcaggagctgcggtacagtgtagccttggtccttaaggagatg 764
          ||||||||||||||||  ||||||||||  |||||  ||  ||  ||||||||||||||||||
Db    319 GTGCCCATAGCCAAATTCCAGGAGCTGCTCTACAGCGTGGCGTTGGTCCTTAAGGAGATG 260

Qy    765 gcagaactggagaagaagtgtgagcgcaaactgcaggactgactgaaccctggtactgt- 824
          |||||  |||||||  ||||||||||||||||||||    ||||||||||||||||||
Db    259 GCAGAGCTGGAGAGAAAGTGTGAGCGCAAACTGCAG----GACTGAACCCTGGTACTGTG 204

Qy    824 gggtgctgaagctggtaccagaacacagccccccactggtgatgagcccaactccattga 883
          ||||||||||||||||||||||||  |||  ||  ||||||||||||
Db    203 GGGTGCTGAAGCTGGTACCAGAACCCAG--CCGCACTGGTGATGA-------------- 160

Qy    884 ggtcctgcatgtgagaacgtattttaagtgaaaagacagc-gggactttcaggttttgtt 942
              |  ||||  |||||  ||||||||||||||||||||  ||||||||  |||||||||
Db    160 ------GGATGTAAGAACATATTTTAAGTGAAAAGACAGCTGGGACTTTTGGGTTTTGTT 107

Qy    943 ttatgagtcaacagctgggcagggtggcacagtttataatctcagcccttggaagtctga 1002
          |||  |||  ||||||||||                                      ||||
Db    106 TTACAAGTAAACAGCTGGG-------------------------------------CTGA 84

Qy   1003 ggctggagaatgggaagtgtaagctgggcctggc------tttcatagtgaggctcagtg 1056
          ||||||||||||  |||  ||||||   ||||||||      ||  |  ||||||||||||||
Db     83 GGCTGGAGAATGTGAAATGTAAGGCCGGCCTGGCTGGGTTTTACGTAGTGAGGCTCAGTT 24

Qy   1057 tcgaattaaagaggtaaagcaac 1079
          |||||||||||||||||||||||
Db     23 TCGAATTAAAGAGGTAAAGCAAC 1


Search completed: May 14, 2000, 23:17:58
Job time: 16358 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: May 14, 2000, 23:17:58 ; Search time 871.67 Seconds
(without alignments)
6300.684 Million cell updates/sec

Title: US-09-223-796-3
Perfect score: 1355
Sequence: 1 gggcaggcagttgaggtgga.....gtgtttcaggcagggcccgg 1355

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: gb_gss1:*
83: gb_gss2:*
84: gb_gss3:*
85: gb_gss4:*
86: em_gss1:*
87: em_gss2:*
88: em_gss3:*
89: em_gss4:*
90: gb_gss5:*
91: gb_gss6:*
92: gb_gss7:*
93: gb_gss8:*
94: gb_gss9:*
95: em_gss5:*
96: em_gss6:*
97: em_gss7:*
98: em_gss8:*
99: em_gss9:*
100: em_gss10:*
101: em_gss11:*
102: gb_gss10:*
103: gb_gss11:*
104: em_gss12:*
105: gb_gss12:*
106: gb_gss13:*
107: gb_gss14:*
108: gb_gss15:*
109: gb_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	% Query

	No.	Score	Match	Length	DB	ID	Description
c	1	540.8	39.9	548	38	AA780068	AA780068 zj24h04.s
c	2	510.8	37.7	520	63	AW005581	AW005581 wz87h11.x
c	3	504.4	37.2	540	94	AQ055548	AQ055548 CIT-HSP-2
c	4	503	37.1	525	61	AI871472	AI871472 wl67e01.x
c	5	503	37.1	527	64	AW025527	AW025527 wu97d10.x
c	6	447	33.0	501	90	AQ798652	AQ798652 HS_2238_B
	7	441.2	32.6	537	27	AA029831	AA029831 zk10b06.r
c	8	440.8	32.5	491	26	W93417	W93417 zd95d01.s1
c	9	427.2	31.5	447	39	AA836105	AA836105 od15c01.s
c	10	420.8	31.1	437	44	AI276262	AI276262 ql65f06.x
c	11	412.8	30.5	431	43	AI240879	AI240879 qj94e02.x
c	12	406.8	30.0	413	42	AI076231	AI076231 oy65c09.x
c	13	405.8	29.9	411	79	AW261921	AW261921 xp87f11.x
c	14	405.8	29.9	418	43	AI189612	AI189612 qd32g06.x
c	15	396.8	29.3	410	39	AA857959	AA857959 oe33g04.s
c	16	396	29.2	406	34	AA464962	AA464962 zx80f06.s
c	17	391.8	28.9	398	43	AI200129	AI200129 qf90h05.x
	18	389.6	28.8	466	21	R20463	R20463 yg33h09.r1
c	19	386.2	28.5	418	26	W86177	W86177 zh59g11.s1
c	20	374.8	27.7	380	44	AI274473	AI274473 qv61c09.x
c	21	374.8	27.7	387	41	AI004350	AI004350 ot95b06.x
c	22	366.6	27.1	386	40	AA994284	AA994284 ou50a05.s
c	23	364.8	26.9	424	25	N49076	N49076 yy79b07.s1
c	24	359.8	26.6	460	45	AI336855	AI336855 qx86g08.x
c	25	358.8	26.5	394	37	AA737662	AA737662 nw59d12.s
c	26	358.2	26.4	374	44	AI248784	AI248784 qh67b01.x
	27	353	26.1	464	26	W86176	W86176 zh59g11.r1
c	28	342.8	25.3	383	38	AA811098	AA811098 oa85e03.s
c	29	340.4	25.1	365	64	AW078888	AW078888 xb19c12.x
c	30	337.6	24.9	410	26	W92692	W92692 zh79d09.s1
c	31	335	24.7	481	27	AA029291	AA029291 zk10b06.s
c	32	329.8	24.3	337	47	AI470273	AI470273 tj18g08.x
c	33	325.8	24.0	487	25	N74187	N74187 za76h04.s1
c	34	313	23.1	455	69	AW139555	AW139555 UI-H-BI1-
	35	310.6	22.9	330	32	AA379908	AA379908 EST92801
c	36	304.8	22.5	356	38	AA780382	AA780382 aa67b02.s
	37	303.6	22.4	311	32	AA378898	AA378898 EST91828
c	38	297	21.9	436	73	AW168901	AW168901 xj14h11.x
c	39	295	21.8	420	46	AI400947	AI400947 th25g10.x
c	40	294	21.7	438	79	AW248567	AW248567 2821036.3
	41	274.8	20.3	361	26	W92672	W92672 zh79d09.r1
	42	268.8	19.8	337	81	AW418433	AW418433 56258 MAR
	43	260.8	19.2	300	32	AA337611	AA337611 EST42359
	44	256.8	19.0	351	30	AA251752	AA251752 zs09e12.r
	45	249.6	18.4	456	27	AA034568	AA034568 mi49a11.r

ALIGNMENTS

```
RESULT   1
AA780068/c
LOCUS       AA780068        548 bp    mRNA            EST       05-FEB-1998
DEFINITION  zj24h04.s1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA
            clone IMAGE:451255 3', mRNA sequence.
ACCESSION   AA780068
VERSION     AA780068.1  GI:2839399
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 548)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 9, 1998 this sequence version replaced gi:936306.
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 507.
FEATURES             Location/Qualifiers
     source          1. .548
                     /organism="Homo sapiens"
                     /db_xref="GDB:1387611"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:451255"
                     /clone_lib="Soares_fetal_liver_spleen_1NFLS_S1"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     This is a subtracted version of the original Soares fetal
                     liver spleen 1NFLS library.  1st strand cDNA was primed
                     with a Pac I - oligo(dT) primer [5'
                     AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      129 a    158 c    152 g    108 t      1 others
ORIGIN
```

Query Match 39.9%; Score 540.8; DB 38; Length 548;
Best Local Similarity 99.4%; Pred. No. 3.5e-125;
Matches 542; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy     627 ccagctccaggagctctgcatcccccaagacctggtcggggacttggccagcgtggtatt 686
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     545 CCAGCTCCAGGAGCTCTGCATCCCCCAAGACCTGGTCGGGGACTTGGCCAGCGTGGTATT 486

Qy     687 tgggagccagcggcccctccttgattctgtggcccagcagcaggggcctggctgccgca 746
           |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db     485 TGGGAGCCAGCGGCCCCTCCTTGATTCTGTGGCCCAGCAGCAGGGGGCCTNGCTGCCGCA 426

Qy     747 tgttgctgactttcggtggcgggtggatgtagcaatctccaccagtgccctggctcgctc 806
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     425 TGTTGCTGACTTTCGGTGGCGGGTGGATGTAGCAATCTCCACCAGTGCCCTGGCTCGCTC 366

Qy     807 cctgcagccgagcgtcctgatgcagctgaagctttcagatgggtcagcataccgctttga 866
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     365 CCTGCAGCCGAGCGTCCTGATGCAGCTGAAGCTTTCAGATGGGTCAGCATACCGCTTTGA 306

Qy     867 ggtccccacagccaagttccaggagctgcggtacagcgtggccctggtcctaaaggagat 926
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     305 GGTCCCCACAGCCAAGTTCCAGGAGCTGCGGTACAGCGTGGCCCTGGTCCTAAAGGAGAT 246

Qy     927 ggcagatctggagaagaggtgtgagcgcagactgcaggactgacccctcacttgaccagt 986
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     245 GGCAGATCTGGAGAAGAGGTGTGAGCGCAGACTGCAGGACTGACCCCTCACTTGACCAGT 186

Qy     987 cccattcagatccggcttggacaggcacctgagatggtgccaaagtgcagctgactcttc 1046
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     185 CCCATTCAGATCCGGCTTGGACAGGCACCTGAGATGGTGCCAAAGTGCAGCTGACTCTTC 126

Qy    1047 ccacgacagccctggccttcccatcaggcaggctcttcagtgagtgtttgaacgtaatta 1106
           |||||||||||||| |||||||| ||||||||||||||||||||||||||||||||||||
Db     125 CCACGACAGCCCTGCCCTTCCCATGAGGCAGGCTCTTCAGTGAGTGTTTGAACGTAATTA 66

Qy    1107 tgtagttttctgtttaattgaaaaagagagctatgccttttttttcttttttggaagtaaag 1166
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65 TGTAGTTTTCTGTTTAATTGAAAAAGAGAGCTATGCCTTTTTTTCTTTTTGGAAGTAAAG 6
```

```
Qy   1167 cagct 1171
          |||||
Db      5 CAGCT 1


RESULT   2
AW005581/c
LOCUS       AW005581      520 bp    mRNA            EST       10-SEP-1999
DEFINITION  wz87h11.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2565861 3',
            mRNA sequence.
ACCESSION   AW005581
VERSION     AW005581.1  GI:5854359
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 520)
  AUTHORS   NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGAP), Tumor Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3187080.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
            Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 452.
FEATURES             Location/Qualifiers
     source          1. .520
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2565861"
                     /clone_lib="NCI_CGAP_Brn25"
                     /tissue_type="anaplastic oligodendroglioma"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATAGGTTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library is normalized, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT      124 a    150 c    137 g    109 t
ORIGIN


  Query Match             37.7%;  Score 510.8;  DB 63;  Length 520;
  Best Local Similarity   99.6%;  Pred. No. 1.1e-117;
  Matches  512;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy    662 tcggggacttggccagcgtggtatttgggagccagcggcccctccttgattctgtggccc 721
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    520 TCGGGGACTTGGCCAGCGTGGTATTTGGGAGCCAGCGGCCCCTCCTTGATTCTGTGGCCC 461

Qy    722 agcagcaggggggcctggctgccgcatgttgctgactttcggtggcgggtggatgtagcaa 781
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    460 AGCAGCAGGGGGGCCTGGCTGCCGCATGTTGCTGACTTTCGGTGGCGGGTGGATGTAGCAA 401

Qy    782 tctccaccagtgccctggctcgctccctgcagccgagcgtcctgatgcagctgaagcttt 841
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    400 TCTCCACCAGTGCCCTGGCTCGCTCCCTGCAGCCGAGCGTCCTGATGCAGCTGAAGCTTT 341

Qy    842 cagatgggtcagcataccgcttgaggtccccacagccaagttccaggagctgcggtaca 901
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    340 CAGATGGGTCAGCATACCGCTTTGAGGTCCCCACAGCCAAGTTCCAGGAGCTGCGGTACA 281

Qy    902 gcgtggccctggtcctaaaggagatggcagatctggagaagaggtgtgagcgcagactgc 961
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    280 GCGTGGCCCTGGTCCTAAAGGAGATGGCAGATCTGGAGAAGAGGTGTGAGCGCAGACTGC 221

Qy    962 aggactgacccctcacttgaccagtcccattcagatccggcttggacaggcacctgagat 1021
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    220 AGGACTGACCCCTCACTTGACCAGTCCCATTCAGATCCGGCTTGGACAGGCACCTGAGAT 161

Qy   1022 ggtgccaaagtgcagctgactcttcccacgacagccctggccttcccatcaggcaggctc 1081
          |||||||||||||||||||||||||||||||||| ||||||||| ||||||||||
Db    160 GGTGCCAAAGTGCAGCTGACTCTTCCCACGACAGCCCTGCCCTTCCCATGAGGCAGGCTC 101

Qy   1082 ttcagtgagtgtttgaacgtaattatgtagttttctgtttaattgaaaaagagagctatg 1141
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    100 TTCAGTGAGTGTTTGAACGTAATTATGTAGTTTTCTGTTTAATTGAAAAAGAGAGCTATG 41

Qy   1142 ccttttttttcttttttggaagtaaagcagctaaaa 1175
          |||||||||||||||||||||||||||||||||||
Db     40 CCTTTTTTTTCTTTTTGGAAGTAAAGCAGCTAAAA 7


RESULT   3
AQ055548/c
LOCUS       AQ055548      540 bp    DNA             GSS       30-JUL-1998
DEFINITION  CIT-HSP-2338P5.TR CIT-HSP Homo sapiens genomic clone 2338P5,
            genomic survey sequence.
ACCESSION   AQ055548
VERSION     AQ055548.1  GI:3352154
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 540)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
            Simon,M. and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building (1998)
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: CIT-HSP-2338P5.TF
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .540
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="2338P5"
                     /clone_lib="CIT-HSP"
                     /sex="Male"
                     /cell_type="Sperm"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
BASE COUNT      107 a    170 c    166 g     97 t
ORIGIN
```

```
  Query Match             37.2%;  Score 504.4;  DB 94;  Length 540;
  Best Local Similarity   98.1%;  Pred. No. 4.6e-116;
  Matches  522;  Conservative    0;  Mismatches    6;  Indels    4;  Gaps    1;

Qy     313 gcaactccatacctgcatcatcctggtgatagtcacagtggccgagtgagtttcttgggg 372
           ||||||||||||| ||||||||||||||| ||||||||||||||||||||||||||||||
Db     540 GCAACTCCATACATGCATCATCCTGGTAATAGTCACAGTGGCCGAGTGAGTTTCTTGGGG 481

Qy     373 gcccagctt----cctccagaggtggcagcaatggcccggctactaggggacctagacag 428
           |||||||||    ||||||||||||||||||||||||||||||||||||||||||||||
Db     480 GCCCAGCTTTCCCACTCCAGAGGTGGCAGCAATGGCCCGGCTACTAGGGGACCTAGACAG 421

Qy     429 gagcacgttcagaaagttgctgaagtttgtggtcagcagcctgcaggggaggactgccg 488
           |||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db     420 GAGCACGTTCAGAAAGTTGCTGAAGTTTGTGGTCAGCAGCCTGCAGGGGAAGGACTGCCG 361

Qy     489 agacggtgtgcagcgtcttggggtcagcgccaacctgccggaggagcagctgggtgccct 548
           ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     360 AGAGGCTGTGCAGCGTCTTGGGGTCAGCGCCAACCTGCCGGAGGAGCAGCTGGGTGCCCT 301

Qy     549 gctggcaggcatgcacacactgctccagcaggccctccgtctgccccccaccagcctgaa 608
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     300 GCTGGCAGGCATGCACACACTGCTCCAGCAGGCCCTCCGTCTGCCCCCCACCAGCCTGAA 241

Qy     609 gcctgacaccttcagggaccagctccaggagctctgcatcccccaagacctggtcgggga 668
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     240 GCCTGACACCTTCAGGGACCAGCTCCAGGAGCTCTGCATCCCCCAAGACCTGGTCGGGGA 181

Qy     669 cttggccagcgtggtatttgggagccagcggcccctccttgattctgtggcccagcagca 728
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180 CTTGGCCAGCGTGGTATTTGGGAGCCAGCGGCCCCTCCTTGATTCTGTGGCCCAGCAGCA 121

Qy     729 gggggcctggctgccgcatgttgctgactttcggtggcgggtggatgtagcaatctccac 788
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     120 GGGGGCCTGGCTGCCGCATGTTGCTGACTTTCGGTGGCGGGTGGATGTAGCAATCTCCAC 61

Qy     789 cagtgccctggctcgctccctgcagccgagcgtcctgatgcagctgaagctt 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60 CAGTGCCCTGGCTCGCTCCCTGCAGCCGAGCGTCCTGATGCAGCTGAAGCTT 9


RESULT   4
AI871472/c
LOCUS       AI871472      525 bp    mRNA            EST       30-AUG-1999
DEFINITION  wl67e01.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2429976 3',
            mRNA sequence.
ACCESSION   AI871472
VERSION     AI871472.1  GI:5545521
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 525)
  AUTHORS   NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGAP), Tumor Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     On Dec 20, 1995 this sequence version replaced gi:1131323.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
            Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40UP from Gibco
            High quality sequence stop: 453.
FEATURES             Location/Qualifiers
     source          1. .525
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2429976"
                     /clone_lib="NCI_CGAP_Brn25"
                     /tissue_type="anaplastic oligodendroglioma"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATAGGTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library is normalized, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT      126 a    152 c    140 g    107 t
ORIGIN


  Query Match             37.1%;  Score 503;  DB 61;  Length 525;
  Best Local Similarity   98.9%;  Pred. No. 1e-115;
  Matches  517;  Conservative    0;  Mismatches    5;  Indels    1;  Gaps    1;

Qy     654 agacctggtcggggacttggccagcgtggtatttgggagccagcggcccctccttgattc 713
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     525 AGACCTGGTCGGGGACTTGGCCAGCGTGGTATTTGGGAGCCAGCGGCCCCTCCTTGATTC 466

Qy     714 tgtggcccagcagcaggggggcctggctgccgcatgttgctgactttcggtggcgggtgga 773
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     465 TGTGGCCCAGCAGCAGGGGGCCTGGCTGCCGCATGTTGCTGACTTTCGGTGGCGGGTGGA 406

Qy     774 tgtagcaatctccaccagtgccctggctcgctccctgcagccgagcgtcctgatgcagct 833
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     405 TGTAGCAATCTCCACCAGTGCCCTGGCTCGCTCCCTGCAGCCGAGCGTCCTGATGCAGCT 346

Qy     834 gaagctttcagatgggtcagcataccgctttgaggtccccacagccaagttccaggagct 893
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     345 GAAGCTTTCAGATGGGTCAGCATACCGCTTTGAGGTCCCCACAGCCAAGTTCCAGGAGCT 286

Qy     894 gcggtacagcgtggccctggtcctaaaggagatggcagatctggagaagaggtgtgag-c 952
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db     285 GCGGTACAGCGTGGCCCTGGTCCTAAAGGAGATGGCAGATCTGGAGAAGAGGTGTGAGCC 226

Qy     953 gcagactgcaggactgacccctcacttgaccagtcccattcagatccggcttggacaggc 1012
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     225 GCAACTTGCAGGACTGACCCCTCACTTGACCAGTCCCATTCAGATCCGGCTTGGACAGGC 166

Qy    1013 acctgagatggtgccaaagtgcagctgactcttcccacgacagccctggccttcccatca 1072
           |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||| |
Db     165 ACCTGAGATGGTGCCAAAGTGCAGCTGACTCTTCCCACGACAGCCCTGCCCTTCCCATGA 106

Qy    1073 ggcaggctcttcagtgagtgtttgaacgtaattatgtagttttctgtttaattgaaaaag 1132
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     105 GGCAGGCTCTTCAGTGAGTGTTTGAACGTAATTATGTAGTTTTCTGTTTAATTGAAAAAG 46

Qy    1133 agagctatgccttttttttcttttttggaagtaaagcagctaaaa 1175
           |||||||||||||||||||||||||||||||||||||||||||||
Db      45 AGAGCTATGCCTTTTTTTCTTTTTGGAAGTAAAGCAGCTAAAA 3


RESULT   5
AW025527/c
LOCUS       AW025527      527 bp    mRNA            EST       27-OCT-1999
DEFINITION  wu97d10.x1 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:2527987 3',
            mRNA sequence.
ACCESSION   AW025527
VERSION     AW025527.1  GI:5879057
```

```
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 527)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3187612.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 444.
FEATURES             Location/Qualifiers
     source          1. .527
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2527987"
                     /clone_lib="NCI_CGAP_Kid3"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer,
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia),  digested with Not I and cloned into the Not
                     I and Eco RI sites  of the modified pT7T3 vector.  mRNA
                     source: 2 pooled kidneys. Library went through one round
                     of normalization. Library constructed by Bento Soares and
                     M. Fatima Bonaldo.  "
BASE COUNT      126 a    150 c    139 g    112 t
ORIGIN
```

```
  Query Match           37.1%;  Score 503;  DB 64;  Length 527;
  Best Local Similarity   98.9%;  Pred. No. 1e-115;
  Matches  517;  Conservative    0;  Mismatches    5;  Indels    1;  Gaps    1;

Qy    654 agacctggtcggggacttggccagcgtggtatttgggagccagcggcccctccttgattc 713
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    527 AGACCTGGTCGGGGACTTGGCCAGCGTGGTATTTGGGAGCCAGCGGCCCCTCCTTGATTC 468

Qy    714 tgtggcccagcagcaggggggcctggctgccgcatgttgctgactttcggtggcgggtgga 773
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    467 TGTGGCCCAGCAGCAGGGGGCCTGGCTGCCGCATGTTGCTGACTTTCGGTGGCGGGTGGA 408

Qy    774 tgtagcaatctccaccagtgccctggctcgctccctgcagccgagcgtcctgatgcagct 833
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    407 TGTAGCAATCTCCACCAGTGCCCTGGCTCGCTCCCTGCAGCCGAGCGTCCTGATGCAGCT 348

Qy    834 gaagctttcagatgggtcagcataccgctttgaggtccccacagccaagttccaggagct 893
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    347 GAAGCTTTCAGATGGGTCAGCATACCGCTTTGAGGTCCCCACAGCCAAGTTCCAGGAGCT 288

Qy    894 gcggtacagcgtggccctggtcctaaaggagatggcagatctgg-agaagaggtgtgagc 952
          |||||||||||||||||||||||||||||||||||||||||| | || || |||||||
Db    287 GCGGTACAGCGTGGCCCTGGTCCTAAAGGAGATGGCAGATCTGGAAAAAAAGTTGTGAGC 228

Qy    953 gcagactgcaggactgacccctcacttgaccagtcccattcagatccggcttggacaggc 1012
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    227 GCAGACTGCAGGACTGACCCCTCACTTGACCAGTCCCATTCAGATCCGGCTTGGACAGGC 168
```

```
Qy   1013 acctgagatggtgccaaagtgcagctgactcttcccacgacagccctggccttcccatca 1072
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||| |
Db    167 ACCTGAGATGGTGCCAAAGTGCAGCTGACTCTTCCCACGACAGCCCTGCCCTTCCCATGA 108

Qy   1073 ggcaggctcttcagtgagtgtttgaacgtaattatgtagttttctgtttaattgaaaaag 1132
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    107 GGCAGGCTCTTCAGTGAGTGTTTGAACGTAATTATGTAGTTTTCTGTTTAATTGAAAAAG 48

Qy   1133 agagctatgccttttttttcttttttggaagtaaagcagctaaaa 1175
          |||||||||||||||||||||||||||||||||||||||||||||
Db     47 AGAGCTATGCCTTTTTTTCTTTTTGGAAGTAAAGCAGCTAAAA 5
```

```
RESULT   6
AQ798652/c
LOCUS       AQ798652      501 bp    DNA             GSS       09-AUG-1999
DEFINITION  HS_2238_B1_H03_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2238 Col=5 Row=P, genomic survey
            sequence.
ACCESSION   AQ798652
VERSION     AQ798652.1  GI:5715984
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 501)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     On Sep 10, 1998 this sequence version replaced gi:3553660.
            Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 2238  row: P  column: 5
            Seq primer: M13 Reverse
            Class: BAC ends
            High quality sequence stop: 501.
FEATURES             Location/Qualifiers
     source          1. .501
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=2238 Col=5 Row=P"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      105 a    163 c    139 g     92 t      2 others
ORIGIN
```

```
  Query Match           33.0%;  Score 447;  DB 90;  Length 501;
  Best Local Similarity   94.4%;  Pred. No. 1.1e-101;
  Matches  473;  Conservative    0;  Mismatches   27;  Indels    1;  Gaps    1;

Qy    334 cctggtgatagtcacagtggccgagtgagtttcttgggggcccagcttcctccagaggtg 393
          | ||||||||||||||||||||||||| |||||||||||  ||||| ||  ||||||
Db    500 CGTGGTGATAGTCACAGTGGCCGAGTGAG-TTCTTGGGGGCCAGGCTTCTTCAGGAGGTG 442

Qy    394 gcagcaatggcccggctactaggggacctagacaggagcacgttcagaaagttgctgaag 453
          |||||||||| |||||||||| || ||||||| ||||||||||||||||||||||||||
Db    441 ACAGCAATGGCCTGGCTACTAGGGAACTTAGACAGAAGCACGTTCAGAAAGTTGCTGAAG 382
```

```
Qy    454 tttgtggtcagcagcctgcaggggggaggactgccgagacggtgtgcagcgtcttggggtc 513
          ||||||||||||||| |||||||| || ||||||||||   ||||||| ||||||||||
Db    381 TTTGTGGTCAGCAGCTTGCAGGGGAAGAACTGCCGAGAGTCTGTGCAGTGTCTTGGGGTC 322

Qy    514 agcgccaacctgccggaggagcagctgggtgccctgctggcaggcatgcacacactgctc 573
          ||||||||| ||||| |||||||||||||||| || ||||||||||||||||||||||||
Db    321 AGCGCCAACATGCCGAAGGAGCAGCTGGGTGCCNTGTTGGCAGGCATGCACACACTGCTC 262

Qy    574 cagcaggccctccgtctgccccccaccagcctgaagcctgacaccttcagggaccagctc 633
          ||||||||||| |||||||||||||||| |||||| ||||||||||||||||||||||||
Db    261 CAGCAGGCCCTCTGTCTGCCCCCCACCAGCNTGAAGCTTGACACCTTCAGGGACCAGCTC 202

Qy    634 caggagctctgcatcccccaagacctggtcggggacttggccagcgtggtatttgggagc 693
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 CAGGAGCTCTGCATCCCCCAAGACCTGGTCGGGGACTTGGCCAGCGTGGTATTTGGGAGC 142

Qy    694 cagcggcccctccttgattctgtggcccagcagcaggggggcctggctgccgcatgttgct 753
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 CAGCGGCCCCTCCTTGATTCTGTGGCCCAGCAGCAGGGGGCCTGGCTGCCGCATGTTGCT 82

Qy    754 gactttcggtggcgggtggatgtagcaatctccaccagtgccctggctcgctccctgcag 813
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db     81 GACTTTCGGTGGCGGGTGGATGTAGCAATCTCCACCAGTGCCATGGCTCGCTCCCTGCAG 22

Qy    814 ccgagcgtcctgatgcagctg 834
          ||||||||| ||||||||||
Db     21 CCGAGCGTCTTGATGCAGCTG 1
```

RESULT 7
AA029831
LOCUS AA029831 537 bp mRNA EST 09-MAY-1997
DEFINITION zk10b06.r1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone IMAGE:470099 5', mRNA sequence.
ACCESSION AA029831
VERSION AA029831.1 GI:1496076
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 537)
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B., Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
MEDLINE 97044478
COMMENT On Apr 14, 1993 this sequence version replaced gi:837792.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1172 Std Error: 0.00
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 348.
FEATURES Location/Qualifiers
source 1. .537
/organism="Homo sapiens"
/db_xref="GDB:3756101"
/db_xref="taxon:9606"
/clone="IMAGE:470099"
/clone_lib="Soares_pregnant_uterus_NbHPU"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' AACTGGAAGAATTCGCGGCCGCCTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by M. Fatima Bonaldo."
BASE COUNT 93 a 159 c 173 g 109 t 3 others
ORIGIN

Query Match 32.6%; Score 441.2; DB 27; Length 537;
Best Local Similarity 96.5%; Pred. No. 3.1e-100;
Matches 524; Conservative 0; Mismatches 11; Indels 8; Gaps 7;

```
Qy    335 ctggtgatagtcacagtggccgagtgagtttcttgggggcccagcttcctccagaggtgg 394
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CTGGTGATAGTCACAGTGGCCGAGTGAGTTTCTTGGGGGCCCAGCTTCCTCCAGAGGTGG 60

Qy    395 cagcaatggcccggctactaggggacctagacaggagcacgttcagaaagttgctgaagt 454
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CAGCAATGGCCCGGCTACTAGGGGACCTAGACAGGAGCACGTTCAGAAAGTTGCTGAAGT 120

Qy    455 ttgtggtcagcagcctgcaggggggaggactgccgagacggtgtgcagcgtcttggggtca 514
          ||||||||||||||||||||||||||||||||||  || ||||||||||||||||||
Db    121 TTGTGGTCAGCAGCCTGCAGGGGGAGGACTGCCGAGANTGT-TGCAGCGTCTTGGGGTCA 179

Qy    515 gcgccaacctgccggaggagcagctgggtgccctgctggcaggcatgcacacactgctcc 574
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    180 GCGCCAACCTGCCGGA-GAGCAGCTGGGTGCCCTGCTGGCAGGCATGCACACACTGCTCC 238

Qy    575 agcaggccctccgtctgccccccaccagcctgaagcctgacaccttcagggaccagctcc 634
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    239 AGCAGGCCCTCCGTCTGCCCCCCACCAGCCTGAAGCCTGACACCTTCAGGGACCAGCTCC 298

Qy    635 aggagctctgcatcccccaagacctggtcggggacttggccagcgtggtatttgggagcc 694
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    299 AGGAGCTCTGCATCCCCCAAGACCTGGTCGGGGACTTGGCCAGCGTGGTATTTGGGAGCC 358

Qy    695 agcggcccctccttgattctgtggcccagcagcaggggggcctggctgccgcatgttgct- 754
          ||||| |||||||||||||||| |||||||||||||| ||||||||||||||||||||
Db    359 AGCGG-CCCTCCTTGATTCTGTGG-CCAGCAGCAGGGGG-CTGGCTGCCGCATGTTGCTT 415

Qy    754 -gactttcggtggcgggtggatgtagcaatctccaccagtgccctggctcgctccctgca 812
           ||  | ||||||||||||||| ||||||||||||||||||||||||||||| ||||||
Db    416 GACTTTCGGTTGGCGGGTGGATGTAACAATCTCCACCAGTGCCCTGGCTCGCTNCCTGCA 475

Qy    813 gccgagcgtcctgatgcagctgaagctttcagatgggtcagcataccgctttgaggtccc 872
          ||||||||| |||||||||||||||||||||||||||||| ||||||||||||||||||
Db    476 GCCGAGCGTNCTGATGCAGCTGAAGCTTTCAGATGGGTCAG-ATACCGCTTTGAGGTCCC 534

Qy    873 cac 875
          |||
Db    535 CAC 537
```

RESULT 8
W93417/c
LOCUS W93417 491 bp mRNA EST 16-JUL-1996
DEFINITION zd95d01.s1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone IMAGE:357217 3', mRNA sequence.
ACCESSION W93417
VERSION W93417.1 GI:1422688
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 491)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,

```
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     On Sep 21, 1992 this sequence version replaced gi:279425.
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: mob.REGA+ET
            High quality sequence stop: 375.
FEATURES             Location/Qualifiers
     source          1. .491
                     /organism="Homo sapiens"
                     /db_xref="GDB:1273761"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:357217"
                     /clone_lib="Soares_fetal_heart_NbHH19W"
                     /sex="unknown"
                     /dev_stage="19 weeks"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATCTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by
                     M.Fatima Bonaldo. This library was constructed from the
                     same fetus as the fetal lung library, Soares fetal lung
                     NbHL19W."
BASE COUNT      116 a    135 c    134 g    103 t      3 others
ORIGIN

  Query Match             32.5%;  Score 440.8;  DB 26;  Length 491;
  Best Local Similarity   97.3%;  Pred. No. 3.8e-100;
  Matches  468;  Conservative    0;  Mismatches   10;  Indels    3;  Gaps    2;

Qy     698 ggcccctccttgattctgtggcccagc--agcaggggcctggctgccgcatgttgctga 755
           | ||| |||||||||||||| ||||||    |||||||||||| ||||||||||||||||
Db     484 GCCCCNTCCTTGATTCTGTGNCCCAGCCAGCCAGGGGGCCTGGGTGCCGCATGTTGCTGA 425

Qy     756 ctttcggtggcgggtggatgtagcaatctccaccagtgccctggctcgctccctgc-agc 814
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||| 
Db     424 CTTTCGGTGGCGGGTGGATGTAGCAATCTCCACCAGTGCCCTGGCTCGCTCCCTGCAAGC 365

Qy     815 cgagcgtcctgatgcagctgaagctttcagatgggtcagcataccgctttgaggtcccca 874
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     364 CGAGCGTCCTGATGCAGCTGAAGCTTTCAGATGGGTCAGCATACCGCTTTGAGGTCCCCA 305

Qy     875 cagccaagttccaggagctgcggtacagcgtggccctggtcctaaaggagatggcagatc 934
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     304 CAGCCAAGTTCCAGGAGCTGCGGTACAGCGTGGCCCTGGTCCTAAAGGAGATGGCAGATC 245

Qy     935 tggagaagaggtgtgagcgcagactgcaggactgacccctcacttgaccagtcccattca 994
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     244 TGGAGAAGAGGTGTGAGCGCAGACTGCAGGACTGACCCCTCACTTGACCAGTCCCATTCA 185

Qy     995 gatccggcttggacaggcacctgagatggtgccaaagtgcagctgactcttcccacgaca 1054
           |||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db     184 GATCCGGCTTGGACNAGCACCTGAGATGGTGCCAAAGTGCAGCTGACTCTTCCCACGACA 125

Qy    1055 gccctggccttcccatcaggcaggctcttcagtgagtgtttgaacgtaattatgtagttt 1114
           |||||| |||||||||| ||||||||||||||||||||||||||||||||||||||||||

Db     124 GCCCTGCCCTTCCCATGAGGCAGGCTCTTCAGTGAGTGTTTGAACGTAATTATGTAGTTT 65

Qy    1115 tctgtttaattgaaaaagagagctatgccttttttttcttttttggaagtaaagcagctaaa 1174
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64 TCTGTTTAATTGAAAAAGAGAGCTATGCCTTTTTTTCTTTTTGGAAGTAAAGCAGCTAAA 5

Qy    1175 a 1175
           |
Db       4 A 4


RESULT   9
AA836105/c
LOCUS       AA836105      447 bp    mRNA            EST       25-MAR-1998
DEFINITION  od15c01.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1368000,
            mRNA sequence.
ACCESSION   AA836105
VERSION     AA836105.1  GI:2910424
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 447)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On May 5, 1995 this sequence version replaced gi:797662.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 1182   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 441.
FEATURES             Location/Qualifiers
     source          1. .447
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1368000"
                     /clone_lib="NCI_CGAP_GCB1"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTT-
                     3']. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      108 a    119 c    112 g    108 t
ORIGIN

  Query Match             31.5%;  Score 427.2;  DB 39;  Length 447;
  Best Local Similarity   99.3%;  Pred. No. 9.3e-97;
  Matches  429;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;
```

```
Qy    744 gcatgttgctgactttcggtggcgggtggatgtagcaatctccaccagtgccctggctcg 803
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    447 GCATGTTGCTGACTTTCGGTGGCGGGTGGATGTAGCAATCTCCACCAGTGCCCTGGCTCG 388

Qy    804 ctccctgcagccgagcgtcctgatgcagctgaagctttcagatgggtcagcataccgctt 863
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    387 CTCCCTGCAGCCGAGCGTCCTGATGCAGCTGAAGCTTTCAGATGGGTCAGCATACCGCTT 328

Qy    864 tgaggtccccacagccaagttccaggagctgcggtacagcgtggccctggtcctaaagga 923
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    327 TGAGGTCCCCACAGCCAAGTTCCAGGAGCTGCGGTACAGCGTGGCCCTGGTCCTAAAGGA 268

Qy    924 gatggcagatctggagaagaggtgtgagcgcagactgcaggactgacccctcacttgacc 983
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    267 GATGGCAGATCTGGAGAAGAGGTGTGAGCGCAGACTGCAGGACTGACCCCTCACTTGACC 208

Qy    984 agtcccattcagatccggcttggacaggcacctgagatggtgccaaagtgcagctgactc 1043
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    207 AGTCCCATTCAGATCCGGCTTGGACAAGCACCTGAGATGGTGCCAAAGTGCAGCTGACTC 148

Qy   1044 ttcccacgacagccctggccttcccatcaggcaggctcttcagtgagtgtttgaacgtaa 1103
          ||||||||||||||||| |||||||| ||||||||||||||||||||||||||||||||||||||
Db    147 TTCCCACGACAGCCCTGCCCTTCCCATGAGGCAGGCTCTTCAGTGAGTGTTTGAACGTAA 88

Qy   1104 ttatgtagttttctgtttaattgaaaaagagagctatgccttttttttcttttggaagta 1163
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     87 TTATGTAGTTTTCTGTTTAATTGAAAAAGAGAGCTATGCCTTTTTTTCTTTTTGGAAGTA 28

Qy   1164 aagcagctaaaa 1175
          ||||||||||||
Db     27 AAGCAGCTAAAA 16
```

RESULT 10
AI276262/c
LOCUS AI276262 437 bp mRNA EST 29-JAN-1999
DEFINITION q165f06.x1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:1877219 3', mRNA sequence.
ACCESSION AI276262
VERSION AI276262.1 GI:3898536
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 437)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2282464.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1029 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 436.
FEATURES Location/Qualifiers
source 1. .437
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1877219"
/clone_lib="Soares_NhHMPu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (melanocyte 2NbHM, pregnant uterus NbHPU, and fetal heart NbHH19W) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."
BASE COUNT 108 a 120 c 113 g 96 t
ORIGIN

Query Match 31.1%; Score 420.8; DB 44; Length 437;
Best Local Similarity 99.3%; Pred. No. 3.7e-95;
Matches 433; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

```
Qy    740 tgccgcatgttgctgactttcggtggcgggtggatgtagcaatctccaccagtgccctgg 799
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    437 TGCCGCATGTTGCTGACTTTCGGTGGCGGGTGGATGTAGCAATCT-CACCAGTGCCCTGG 379

Qy    800 ctcgctccctgcagccgagcgtcctgatgcagctgaagctttcagatgggtcagcatacc 859
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    378 CTCGCTCCCTGCAGCCGAGCGTCCTGATGCAGCTGAAGCTTTCAGATGGGTCAGCATACC 319

Qy    860 gctttgaggtccccacagccaagttccaggagctgcggtacagcgtggccctggtcctaa 919
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    318 GCTTTGAGGTCCCCACAGCCAAGTTCCAGGAGCTGCGGTACAGCGTGGCCCTGGTCCTAA 259

Qy    920 aggagatggcagatctggagaagaggtgtgagcgcagactgcaggactgacccctcactt 979
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    258 AGGAGATGGCAGATCTGGAGAAGAGGTGTGAGCGCAGACTGCAGGACTGACCCCTCACTT 199

Qy    980 gaccagtcccattcagatccggcttggacaggcacctgagatggtgccaaagtgcagctg 1039
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    198 GACCAGTCCCATTCAGATCCGGCTTGGACAGGCACCTGAGATGGTGCCAAAGTGCAGCTG 139

Qy   1040 actcttcccacgacagccctggccttcccatcaggcaggctcttcagtgagtgtttgaac 1099
          |||||||||||||||||| |||||||| |||||||||||||||||||||||||||||||||||||
Db    138 ACTCTTCCCACGACAGCCCTGCCCTTCCCATGAGGCAGGCTCTTCAGTGAGTGTTTGAAC 79

Qy   1100 gtaattatgtagttttctgtttaattgaaaaagagagctatgccttttttttcttttgga 1159
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     78 GTAATTATGTAGTTTTCTGTTTAATTGAAAAAGAGAGCTATGCCTTTTTTTCTTTTTGGA 19

Qy   1160 agtaaagcagctaaaa 1175
          ||||||||||||||||
Db     18 AGTAAAGCAGCTAAAA 3
```

RESULT 11
AI240879/c
LOCUS AI240879 431 bp mRNA EST 28-JAN-1999
DEFINITION qj94e02.x1 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1867130 3', mRNA sequence.
ACCESSION AI240879
VERSION AI240879.1 GI:3836276
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 431)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center

```
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 1066   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 404.
FEATURES             Location/Qualifiers
     source          1. .431
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1867130"
                     /clone_lib="NCI_CGAP_Kid3"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer,
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not
                     I and Eco RI sites of the modified pT7T3 vector. mRNA
                     source: 2 pooled kidneys. Library went through one round
                     of normalization. Library constructed by Bento Soares and
                     M. Fatima Bonaldo. "
BASE COUNT      100 a    115 c    109 g    107 t
ORIGIN


  Query Match             30.5%;  Score 412.8;  DB 43;  Length 431;
  Best Local Similarity   99.5%;  Pred. No. 3.7e-93;
  Matches  414;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy     760 cggtggcgggtggatgtagcaatctccaccagtgccctggctcgctccctgcagccgagc 819
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     431 CGGTGGCGGGTGGATGTAGCAATCTCCACCAGTGCCCTGGCTCGCTCCCTGCAGCCGAGC 372

Qy     820 gtcctgatgcagctgaagctttcagatgggtcagcataccgctttgaggtccccacagcc 879
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     371 GTCCTGATGCAGCTGAAGCTTTCAGATGGGTCAGCATACCGCTTTGAGGTCCCCACAGCC 312

Qy     880 aagttccaggagctgcggtacagcgtggccctggtcctaaaggagatggcagatctggag 939
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     311 AAGTTCCAGGAGCTGCGGTACAGCGTGGCCCTGGTCCTAAAGGAGATGGCAGATCTGGAG 252

Qy     940 aagaggtgtgagcgcagactgcaggactgacccctcacttgaccagtcccattcagatcc 999
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     251 AAGAGGTGTGAGCGCAGACTGCAGGACTGACCCCTCACTTGACCAGTCCCATTCAGATCC 192

Qy    1000 ggcttggacaggcacctgagatggtgccaaagtgcagctgactcttcccacgacagccct 1059
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     191 GGCTTGGACAGGCACCTGAGATGGTGCCAAAGTGCAGCTGACTCTTCCCACGACAGCCCT 132

Qy    1060 ggccttcccatcaggcaggctcttcagtgagtgtttgaacgtaattatgtagttttctgt 1119
           | |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     131 GCCCTTCCCATGAGGCAGGCTCTTCAGTGAGTGTTTGAACGTAATTATGTAGTTTTCTGT 72

Qy    1120 ttaattgaaaaagagagctatgccttttttttcttttttggaagtaaagcagctaaaa 1175
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71 TTAATTGAAAAAGAGAGCTATGCCTTTTTTTTCTTTTTGGAAGTAAAGCAGCTAAAA 16


RESULT  12
AI076231/c
LOCUS       AI076231      413 bp    mRNA            EST      24-SEP-1998
DEFINITION  oy65c09.x1 NCI_CGAP_CLL1 Homo sapiens cDNA clone IMAGE:1670704 3',
            mRNA sequence.
ACCESSION   AI076231
VERSION     AI076231.1  GI:3405409
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 413)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2152808.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,
            M.D., Louis M. Staudt, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 1034   Std Error: 0.00
            Seq primer: -40ml3 fwd. ET from Amersham
            High quality sequence stop: 397.
FEATURES             Location/Qualifiers
     source          1. .413
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1670704"
                     /clone_lib="NCI_CGAP_CLL1"
                     /tissue_type="B-cell, chronic lymphotic leukemia"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATTGCTTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library is normalized, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT       99 a    111 c    108 g     95 t
ORIGIN


  Query Match             30.0%;  Score 406.8;  DB 42;  Length 413;
  Best Local Similarity   99.5%;  Pred. No. 1.2e-91;
  Matches  408;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy     766 cgggtggatgtagcaatctccaccagtgccctggctcgctccctgcagccgagcgtcctg 825
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     413 CGGGTGGATGTAGCAATCTCCACCAGTGCCCTGGCTCGCTCCCTGCAGCCGAGCGTCCTG 354

Qy     826 atgcagctgaagctttcagatgggtcagcataccgctttgaggtccccacagccaagttc 885
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     353 ATGCAGCTGAAGCTTTCAGATGGGTCAGCATACCGCTTTGAGGTCCCCACAGCCAAGTTC 294

Qy     886 caggagctgcggtacagcgtggccctggtcctaaaggagatggcagatctggagaagagg 945
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     293 CAGGAGCTGCGGTACAGCGTGGCCCTGGTCCTAAAGGAGATGGCAGATCTGGAGAAGAGG 234

Qy     946 tgtgagcgcagactgcaggactgacccctcacttgaccagtcccattcagatccggcttg 1005
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     233 TGTGAGCGCAGACTGCAGGACTGACCCCTCACTTGACCAGTCCCATTCAGATCCGGCTTG 174

Qy    1006 gacaggcacctgagatggtgccaaagtgcagctgactcttcccacgacagccctggcctt 1065
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db     173 GACAGGCACCTGAGATGGTGCCAAAGTGCAGCTGACTCTTCCCACGACAGCCCTGCCCTT 114

Qy    1066 cccatcaggcaggctcttcagtgagtgtttgaacgtaattatgtagttttctgtttaatt 1125
           ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     113 CCCATGAGGCAGGCTCTTCAGTGAGTGTTTGAACGTAATTATGTAGTTTTCTGTTTAATT 54

Qy    1126 gaaaaagagagctatgccttttttttcttttttggaagtaaagcagctaaaa 1175
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      53 GAAAAAGAGAGCTATGCCTTTTTTTTCTTTTTGGAAGTAAAGCAGCTAAAA 4
```

```
RESULT  13
AW261921/c
LOCUS       AW261921      411 bp    mRNA            EST       28-DEC-1999
DEFINITION  xp87f11.x1 NCI_CGAP_Brn53 Homo sapiens cDNA clone IMAGE:2747373 3',
            mRNA sequence.
ACCESSION   AW261921
VERSION     AW261921.1  GI:6638737
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 411)
  AUTHORS   NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGAP), Tumor Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3189186.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life
            Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The
            I.M.A.G.E. Consortium DNA Sequencing by: Washington University
            Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Possible reversed clone: polyT not found
            Seq primer: -40UP from Gibco
            High quality sequence stop: 396.
FEATURES             Location/Qualifiers
     source          1. .411
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2747373"
                     /clone_lib="NCI_CGAP_Brn53"
                     /tissue_type="three pooled meningiomas"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                      Library constructed by Life Technologies."
BASE COUNT       99 a    111 c    107 g     94 t
ORIGIN
```

```
  Query Match           29.9%;  Score 405.8;  DB 79;  Length 411;
  Best Local Similarity   99.5%;  Pred. No. 2.1e-91;
  Matches  407;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy     767 gggtggatgtagcaatctccaccagtgccctggctcgctccctgcagccgagcgtcctga 826
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     411 GGGTGGATGTAGCAATCTCCACCAGTGCCCTGGCTCGCTCCCTGCAGCCGAGCGTCCTGA 352

Qy     827 tgcagctgaagctttcagatgggtcagcataccgctttgaggtccccacagccaagttcc 886
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     351 TGCAGCTGAAGCTTTCAGATGGGTCAGCATACCGCTTTGAGGTCCCCACAGCCAAGTTCC 292

Qy     887 aggagctgcggtacagcgtggccctggtcctaaaggagatggcagatctggagaagaggt 946
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     291 AGGAGCTGCGGTACAGCGTGGCCCTGGTCCTAAAGGAGATGGCAGATCTGGAGAAGAGGT 232

Qy     947 gtgagcgcagactgcaggactgacccctcacttgaccagtcccattcagatccggcttgg 1006
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     231 GTGAGCGCAGACTGCAGGACTGACCCCTCACTTGACCAGTCCCATTCAGATCCGGCTTGG 172

Qy    1007 acaggcacctgagatggtgccaaagtgcagctgactcttcccacgacagccctggccttc 1066
           |||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db     171 ACAGGCACCTGAGATGGTGCCAAAGTGCAGCTGACTCTTCCCACGACAGCCCTGCCCTTC 112

Qy    1067 ccatcaggcaggctcttcagtgagtgtttgaacgtaattatgtagttttctgtttaattg 1126
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     111 CCATGAGGCAGGCTCTTCAGTGAGTGTTTGAACGTAATTATGTAGTTTTCTGTTTAATTG 52

Qy    1127 aaaaagagagctatgccttttttttcttttggaagtaaagcagctaaaa 1175
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db      51 AAAAAGAGAGCTATGCCTTTTTTTCTTTTTGGAAGTAAAGCAGCTAAAA 3
```

```
RESULT  14
AI189612/c
LOCUS       AI189612      418 bp    mRNA            EST       28-OCT-1998
DEFINITION  qd32g06.x1 Soares_placenta_8to9weeks_2NbHP8to9W Homo sapiens cDNA
            clone IMAGE:1725466 3', mRNA sequence.
ACCESSION   AI189612
VERSION     AI189612.1  GI:3740821
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 418)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On May 9, 1995 this sequence version replaced gi:803101.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1048   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 391.
FEATURES             Location/Qualifiers
     source          1. .418
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1725466"
                     /clone_lib="Soares_placenta_8to9weeks_2NbHP8to9W"
                     /dev_stage="two placentae: one from 8 weeks and another
                     from 9 weeks post conception"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGATTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library constructed by Bento Soares and
                     M.Fatima Bonaldo."
BASE COUNT       99 a    111 c    107 g    101 t
ORIGIN
```

```
  Query Match           29.9%;  Score 405.8;  DB 43;  Length 418;
  Best Local Similarity   99.5%;  Pred. No. 2.1e-91;
  Matches  407;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy     767 gggtggatgtagcaatctccaccagtgccctggctcgctccctgcagccgagcgtcctga 826
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     418 GGGTGGATGTAGCAATCTCCACCAGTGCCCTGGCTCGCTCCCTGCAGCCGAGCGTCCTGA 359

Qy     827 tgcagctgaagctttcagatgggtcagcataccgctttgaggtccccacagccaagttcc 886
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     358 TGCAGCTGAAGCTTTCAGATGGGTCAGCATACCGCTTTGAGGTCCCCACAGCCAAGTTCC 299

Qy     887 aggagctgcggtacagcgtggccctggtcctaaaggagatggcagatctggagaagaggt 946
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     298 AGGAGCTGCGGTACAGCGTGGCCCTGGTCCTAAAGGAGATGGCAGATCTGGAGAAGAGGT 239
```

```
Qy     947 gtgagcgcagactgcaggactgacccctcacttgaccagtcccattcagatccggcttgg 1006
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     238 GTGAGCGCAGACTGCAGGACTGACCCCTCACTTGACCAGTCCCATTCAGATCCGGCTTGG 179

Qy    1007 acaggcacctgagatggtgccaaagtgcagctgactcttcccacgacagccctggccttc 1066
           ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db     178 ACAGGCACCTGAGATGGTGCCAAAGTGCAGCTGACTCTTCCCACGACAGCCCTGCCCTTC 119

Qy    1067 ccatcaggcaggctcttcagtgagtgtttgaacgtaattatgtagttttctgtttaattg 1126
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     118 CCATGAGGCAGGCTCTTCAGTGAGTGTTTGAACGTAATTATGTAGTTTTCTGTTTAATTG 59

Qy    1127 aaaaagagagctatgccttttttttctttttggaagtaaagcagctaaaa 1175
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      58 AAAAAGAGAGCTATGCCTTTTTTTCTTTTTGGAAGTAAAGCAGCTAAAA 10
```

RESULT 15
AA857959/c
LOCUS AA857959 410 bp mRNA EST 25-MAR-1998
DEFINITION oe33g04.s1 NCI_CGAP_Pr25 Homo sapiens cDNA clone IMAGE:1410390 3', mRNA sequence.
ACCESSION AA857959
VERSION AA857959.1 GI:2946261
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 410)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2284841.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Suzanne L. Topalian, M.D., Robert K. Bright, Ph.D.
cDNA Library Preparation: Stratagene, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1238 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 401.

FEATURES Location/Qualifiers
source 1. .410
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1410390"
/clone_lib="NCI_CGAP_Pr25"
/tissue_type="epithelium (cell line)"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: prostate; Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Normal prostate epithelial cell line (HPV immortalized). 5' adaptor sequence: 5' GAATTCGGCACGAG 3' 3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

BASE COUNT 97 a 105 c 107 g 101 t
ORIGIN

Query Match 29.3%; Score 396.8; DB 39; Length 410;
Best Local Similarity 99.5%; Pred. No. 3.7e-89;
Matches 398; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy     776 tagcaatctccaccagtgccctggctcgctccctgcagccgagcgtcctgatgcagctga 835
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     410 TAGCAATCTCCACCAGTGCCCTGGCTCGCTCCCTGCAGCCGAGCGTCCTGATGCAGCTGA 351

Qy     836 agctttcagatgggtcagcataccgctttgaggtccccacagccaagttccaggagctgc 895
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     350 AGCTTTCAGATGGGTCAGCATACCGCTTTGAGGTCCCCACAGCCAAGTTCCAGGAGCTGC 291

Qy     896 ggtacagcgtggccctggtcctaaaggagatggcagatctggagaagaggtgtgagcgca 955
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     290 GGTACAGCGTGGCCCTGGTCCTAAAGGAGATGGCAGATCTGGAGAAGAGGTGTGAGCGCA 231

Qy     956 gactgcaggactgacccctcacttgaccagtcccattcagatccggcttggacaggcacc 1015
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     230 GACTGCAGGACTGACCCCTCACTTGACCAGTCCCATTCAGATCCGGCTTGGACAGGCACC 171

Qy    1016 tgagatggtgccaaagtgcagctgactcttcccacgacagccctggccttcccatcaggc 1075
           |||||||||||||||||||||||||||||||||||||||||| ||||||||| ||||
Db     170 TGAGATGGTGCCAAAGTGCAGCTGACTCTTCCCACGACAGCCCTGCCCTTCCCATGAGGC 111

Qy    1076 aggctcttcagtgagtgtttgaacgtaattatgtagttttctgtttaattgaaaaagaga 1135
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     110 AGGCTCTTCAGTGAGTGTTTGAACGTAATTATGTAGTTTTCTGTTTAATTGAAAAAGAGA 51

Qy    1136 gctatgccttttttttctttttggaagtaaagcagctaaaa 1175
           ||||||||||||||||||||||||||||||||||||||||
Db      50 GCTATGCCTTTTTTTCTTTTTGGAAGTAAAGCAGCTAAAA 11
```

Search completed: May 14, 2000, 23:18:06
Job time: 16366 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: May 14, 2000, 22:50:12 ; Search time 76.12 Seconds
(without alignments)
1834.863 Million cell updates/sec

Title: US-09-223-796-1
Perfect score: 1100
Sequence: 1 gcacgagccacagccagcta.....attaaaaaaaaaaaaaaaa 1100

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 226296 seqs, 63486255 residues

Total number of hits satisfying chosen parameters: 452592

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/6_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	45.6	4.1	7218	1	US-08-232-463-14	Sequence 14, Appl
	2	37.4	3.4	6727	5	US-08-629-643A-5	Sequence 5, Appli
	3	37.2	3.4	2757	3	US-08-627-254C-26	Sequence 26, Appl
c	4	36.6	3.3	5687	3	US-08-380-403A-3	Sequence 3, Appli
c	5	36.6	3.3	5687	4	US-08-895-628-3	Sequence 3, Appli
	6	35.6	3.2	2908	5	US-08-487-799-1	Sequence 1, Appli
c	7	35	3.2	9515	2	US-08-920-812-13	Sequence 13, Appl
c	8	35	3.2	9515	2	US-08-920-827-13	Sequence 13, Appl
c	9	35	3.2	9515	2	US-08-921-177-13	Sequence 13, Appl
c	10	35	3.2	9515	2	US-08-362-577C-13	Sequence 13, Appl
c	11	35	3.2	9515	3	US-08-920-828-13	Sequence 13, Appl
c	12	34.8	3.2	17041	1	US-08-076-011-1	Sequence 1, Appli
	13	34.8	3.2	246240	3	US-08-724-394A-20	Sequence 20, Appl
	14	34.8	3.2	246240	3	US-08-724-394A-21	Sequence 21, Appl
	15	34.8	3.2	246240	3	US-08-724-394A-22	Sequence 22, Appl
c	16	34.2	3.1	12847	2	US-08-550-715-1	Sequence 1, Appli
	17	33.6	3.1	3323	4	US-08-422-699A-10	Sequence 10, Appl
	18	33.6	3.1	3323	4	US-08-422-706B-10	Sequence 10, Appl
c	19	33.4	3.0	35060	5	US-08-814-095-7	Sequence 7, Appli
c	20	32.8	3.0	1334	4	US-08-481-658B-44	Sequence 44, Appl
c	21	32.8	3.0	1334	4	US-08-477-504A-44	Sequence 44, Appl
c	22	32.8	3.0	1334	4	US-08-486-756A-44	Sequence 44, Appl
c	23	32.8	3.0	1334	4	US-08-485-862B-44	Sequence 44, Appl
c	24	32.8	3.0	1334	5	US-08-787-739-44	Sequence 44, Appl
	25	32.8	3.0	6306	1	US-08-466-390-3	Sequence 3, Appli
	26	32.8	3.0	6306	2	US-08-470-950-3	Sequence 3, Appli
	27	32.8	3.0	6306	2	US-08-467-781-3	Sequence 3, Appli
	28	32.8	3.0	6306	2	US-08-195-487-3	Sequence 3, Appli
	29	32.8	3.0	6306	3	US-08-483-924-3	Sequence 3, Appli
	30	32.8	3.0	6306	6	PCT-US93-06160-3	Sequence 3, Appli
c	31	32.8	3.0	10898	4	US-08-481-658B-5	Sequence 5, Appli
c	32	32.8	3.0	10898	4	US-08-477-504A-5	Sequence 5, Appli
c	33	32.8	3.0	10898	4	US-08-486-756A-5	Sequence 5, Appli
c	34	32.8	3.0	10898	4	US-08-485-862B-5	Sequence 5, Appli
c	35	32.8	3.0	10898	5	US-08-787-739-5	Sequence 5, Appli
c	36	32.6	3.0	1294	4	US-08-665-647-4	Sequence 4, Appli
	37	32.4	2.9	3957	1	US-07-689-008-5	Sequence 5, Appli
	38	32.4	2.9	9540	1	US-07-689-008-1	Sequence 1, Appli
	39	32	2.9	43795	5	US-08-742-185-101	Sequence 101, App
c	40	31.6	2.9	1548	5	US-08-665-259-28	Sequence 28, Appl
c	41	31.6	2.9	1589	1	US-07-971-092-1	Sequence 1, Appli
c	42	31.6	2.9	7676	2	US-08-451-777A-7	Sequence 7, Appli
c	43	31.6	2.9	7676	3	US-08-451-778A-7	Sequence 7, Appli
c	44	31.6	2.9	7676	3	US-08-998-208-7	Sequence 7, Appli
c	45	31.6	2.9	7676	6	PCT-US95-06743-7	Sequence 7, Appli

ALIGNMENTS

```
RESULT   1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO: 14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-Fls
US-08-232-463-14
```

```
  Query Match             4.1%;  Score 45.6;  DB 1;  Length 7218;
  Best Local Similarity   1.7%;  Pred. No. 0.00087;
  Matches    6;  Conservative  207;  Mismatches  141;  Indels    0;  Gaps    0;

Qy     747 ttggtccttaaggagatggcagaactggagaagaagtgtgagcgcaaactgcaggactga 806
           ||||| |  ::::::: :: ::::  :::::::::: : ::: : :::  : ::::  ::
Db    1441 TTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1382

Qy     807 ctgaaccctggtactgtgggtgctgaagctggtaccagaacacagccccccactggtgat 866
           :::    :: :  : ::: :  ::::  :: :  :::: : ::       :  :: ::
Db    1381 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1322

Qy     867 gagcccaactccattgaggtcctgcatgtgagaacgtattttaagtgaaaagacagcggg 926
           :::   ::    :  ::::    : : : ::::: : :    ::: :::::::: :: :::
Db    1321 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1262

Qy     927 actttcaggttttgttttatgagtcaacagctgggcagggtggcacagtttataatctca 986
           :     :::    :   : :::  :: ::  ::: :::: :: : ::   : ::    :
Db    1261 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1202

Qy     987 gcccttggaagtctgaggctggagaatgggaagtgtaagctgggcctggctttcatagtg 1046
           :     :::::   ::::  :::::: :::::: : :::  :::   ::      : :: :
Db    1201 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1142

Qy    1047 aggctcagtgtcgaattaaagaggtaaagcaactattaaaaaaaaaaaaaaaa 1100
           :::   :: :  :::  ::::::: :::: ::  :  :::::::::::::::::
Db    1141 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1088


RESULT   2
US-08-629-643A-5
; Sequence 5, Application US/08629643A
; Patent No. 6025539
;  GENERAL INFORMATION:
;    APPLICANT:  Lee, J. L.
;    APPLICANT:  Lee, N. A.
;    TITLE OF INVENTION:  IL-5 TRANSGENIC MOUSE
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schwegman, Lundberg, Woessner & Kluth, P. A.
;      STREET:  P. O. Box 2938
;      CITY:  Minneapolis
;      STATE:  MN
;      COUNTRY:  USA
;      ZIP:  55402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/629,643A
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/629,643
;      FILING DATE:  09-APR-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Viksnins, Ann S
;      REGISTRATION NUMBER:  37,748
;      REFERENCE/DOCKET NUMBER:  150.167WO1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  612-359-3260
;      TELEFAX:  612-359-3263
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6727 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  650...3771
;      OTHER INFORMATION:  Join 650..730, 1560..1592,
;      OTHER INFORMATION:  3468..3596, 3676..3771
US-08-629-643A-5


  Query Match             3.4%;  Score 37.4;  DB 5;  Length 6727;
  Best Local Similarity   50.9%;  Pred. No. 0.22;
  Matches   89;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy     921 agcgggactttcaggttttgttttatgagtcaacagctgggcagggtggcacagtttata 980
           ||  ||  |  | ||||   ||||  | | || ||     || ||||||  | | | ||
Db    1844 AGGTAGATATGAAAGTTTGTGCTTATAAATAAAAAGAATATGAGAGTGGCAAATTATGTA 1903

Qy     981 atctcagcccttggaagtctgaggctggagaatgggaagtgtaagctgggcctggctttc 1040
           ||| |||  ||||| || |  |||| || | |     ||| || |    || | | |
Db    1904 ATCCCAGTACTTGGGAGCCAAAGGCAGGGGTAGTCTGAGTCTAGGGCCAGCTTAGATACA 1963

Qy    1041 atagtgaggctcagtgtcgaattaaagaggtaaagcaactattaaaaaaaaaaaa 1095
           |     |  |    | | ||    ||||  ||  |  ||||| |  | ||
Db    1964 TTGCCCTGTATGTATCAAAAGTAAATCCTATAAATAAATAAACAAAAACATTAGA 2018


RESULT   3
US-08-627-254C-26
; Sequence 26, Application US/08627254C
; Patent No. 5859229
;  GENERAL INFORMATION:
;    APPLICANT:  Kniss, Douglas A.
;    TITLE OF INVENTION:  Eicosanoid Formation
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Calfee, Halter & Griwold LLP
;      STREET:  800 Superior Avenue
;      CITY:  Cleveland
;      STATE:  Ohio
;      COUNTRY:  USA
;      ZIP:  44114
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/627,254C
;      FILING DATE:
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Golrick, Mary E
;      REGISTRATION NUMBER:  34,829
;      REFERENCE/DOCKET NUMBER:  18525/00107
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (216) 622-8200
;      TELEFAX:  (216) 241-0816
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2757 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-627-254C-26


  Query Match             3.4%;  Score 37.2;  DB 3;  Length 2757;
  Best Local Similarity   54.5%;  Pred. No. 0.17;
  Matches   97;  Conservative    0;  Mismatches   78;  Indels    3;  Gaps    1;

Qy     925 ggactttcaggttttgttttatgagtcaacagctgggcagggtggcacagtttataatct 984
           ||| | |  | ||     ||   | |  |  |  |||   ||| |||||     |||||
```

```
Db   2110 GGAATATAGGCTTAAAACTTTATATTATAGGGTAGGGTGTGGTTGCACACACCTTAATCC 2169

Qy    985 cagcccttggaagtctgaggc---tggagaatgggaagtgtaagctgggcctggctttca 1041
          |||| |||||||| | |||||   ||||     || ||| | ||    |  |||| |  |
Db   2170 CAGCACTTGGAAGGCAGAGGCAGTTGGATCTCTGGGAGTTTGAGGCCAGTTTGGCCTATA 2229

Qy   1042 tagtgaggctcagtgtcgaattaaagaggtaaagcaactattaaaaaaaaaaaaaaaa 1099
          |||||||    ||    |   |  |    ||  |  ||| ||    |||| ||| |||
Db   2230 TAGTGAGTTCTAGGCCAGCCATGGATGCATAGTGAGACTCTTTCTCAAAACAAACAAA 2287


RESULT   4
US-08-380-403A-3/c
; Sequence 3, Application US/08380403A
; Patent No. 5831024
;  GENERAL INFORMATION:
;    APPLICANT:  MINATO, Nagahiro
;    APPLICANT:  HATTORI, Masakazu
;    APPLICANT:  HIROSHI, Kubota
;    APPLICANT:  MASATSUGU, Maeda
;    TITLE OF INVENTION:  SPA-1 PROTEIN AND GENE CODING THEREFOR
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/380,403A
;      FILING DATE:  30-JAN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/325,909
;      FILING DATE:  19-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 6-279712
;      FILING DATE:  20-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 6-139513
;      FILING DATE:  30-MAY-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  WEGNER, Harold C.
;      REGISTRATION NUMBER:  25,258
;      REFERENCE/DOCKET NUMBER:  53466/128/AAOK
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5687 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-380-403A-3


  Query Match             3.3%;  Score 36.6;  DB 3;  Length 5687;
  Best Local Similarity   56.1%;  Pred. No. 0.35;
  Matches   69;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy    977 tataatctcagcccttggaagtctgaggctggagaatgggaagtgtaagctgggcctggc 1036
          | |||||  ||| || || | || ||||| ||   ||   |||  |||   ||||||
Db   2712 TTTAATCCTAGCACTGGGGAATCAGAGGCAGGCAGATTTCGAGTTCAAGGCCAGCCTGGT 2653

Qy   1037 tttcatagtgaggctcagtgtcgaattaaagaggtaaagcaactattaaaaaaaaaaaaa 1096
           | || ||||||   |||  | |        |   | || |||  |      ||||| ||
Db   2652 CTACAGAGTGAGTTCCAGGATAGCCAGGGCTACACAGAGAAACCCTGTCTTGAAAAACAA 2593

Qy   1097 aaa 1099
          |||
Db   2592 AAA 2590


RESULT   5
US-08-895-628-3/c
; Sequence 3, Application US/08895628
; Patent No. 5998585
;  GENERAL INFORMATION:
;    APPLICANT:  MINATO, Nagahiro
;    APPLICANT:  HATTORI, Masakazu
;    APPLICANT:  HIROSHI, Kubota
;    APPLICANT:  MASATSUGU, Maeda
;    TITLE OF INVENTION:  SPA-1 PROTEIN AND GENE CODING THEREFOR
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/895,628
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/380,403
;      FILING DATE:  30-JAN-1995
;      APPLICATION NUMBER:  US 08/325,909
;      FILING DATE:  19-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 6-279712
;      FILING DATE:  20-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 6-139513
;      FILING DATE:  30-MAY-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  WEGNER, Harold C.
;      REGISTRATION NUMBER:  25,258
;      REFERENCE/DOCKET NUMBER:  53466/128/AAOK
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5687 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-895-628-3


  Query Match             3.3%;  Score 36.6;  DB 4;  Length 5687;
  Best Local Similarity   56.1%;  Pred. No. 0.35;
  Matches   69;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy    977 tataatctcagcccttggaagtctgaggctggagaatgggaagtgtaagctgggcctggc 1036
          | |||||  ||| || || | || ||||| ||   ||   |||  |||   ||||||
Db   2712 TTTAATCCTAGCACTGGGGAATCAGAGGCAGGCAGATTTCGAGTTCAAGGCCAGCCTGGT 2653
```

```
Qy   1037 tttcatagtgaggctcagtgtcgaattaaagaggtaaagcaactattaaaaaaaaaaaaa 1096
          | || ||||||   |||  | |        |   | || |||  |      ||||| ||
Db   2652 CTACAGAGTGAGTTCCAGGATAGCCAGGGCTACACAGAGAAACCCTGTCTTGAAAAACAA 2593

Qy   1097 aaa 1099
          |||
Db   2592 AAA 2590


RESULT   6
US-08-487-799-1
; Sequence 1, Application US/08487799C
; Patent No. 6010908
; GENERAL INFORMATION:
;  APPLICANT: Gruenert, Deiter C.
;  APPLICANT: Kunzelmann, Karl
;  TITLE OF INVENTION: GENE THERAPY BY SMALL FRAGMENTS HOMOLOGOUS REPLACEMENT
;  FILE REFERENCE: 480.18-1(HV)
;  CURRENT APPLICATION NUMBER: US/08/487,799C
;  CURRENT FILING DATE: 1995-06-07
;  EARLIER APPLICATION NUMBER: 07/933,471
;  EARLIER FILING DATE: 1992-08-21
;  EARLIER APPLICATION NUMBER: 08/409,544
;  EARLIER FILING DATE: 1995-03-24
;  NUMBER OF SEQ ID NOS: 87
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 2908
;   TYPE: DNA
;   ORGANISM: human
US-08-487-799-1


  Query Match             3.2%;  Score 35.6;  DB 5;  Length 2908;
  Best Local Similarity   71.2%;  Pred. No. 0.51;
  Matches   47;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy   1030 gcctggctttcatagtgaggctcagtgtcgaattaaagaggtaaagcaactattaaaaaa 1089
          |||||| |  || |||||| ||| || || || |||| |  ||||  ||  | |||| ||
Db    126 gcctgggtgacagagtgagactctgtctctaaataaataaataaataaataaataataa 185

Qy   1090 aaaaaa 1095
          | ||||
Db    186 ataaaa 191


RESULT   7
US-08-920-812-13/c
: Sequence 13, Application US/08920812
; Patent No. 5763188
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/920,812
;      FILING DATE:  29-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/362,577
;      FILING DATE:  27-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9515 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Pseudomonas aeruginosa
;      STRAIN:  Clinical Isolate P2-2
US-08-920-812-13


  Query Match             3.2%;  Score 35;  DB 2;  Length 9515;
  Best Local Similarity   56.5%;  Pred. No. 1.3;
  Matches   65;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy    318 gactgcagagaagctgtggagcaacttggtgccagcgccaacctgtcagaagagcgtctg 377
          |||  | | | |||  | ||||| |   | | |   |  |     ||  ||  ||| | ||
Db   1222 GACCTCGGTGGAGCCGGTGAACACCAGTGCGTCCACCTCCGGGTGCAAGCCGAGGGCCTT 1163

Qy    378 gccgtcctgctggcgggcacacacaccctgctccagcaggctctccggctgcccc 432
          |||| |||||| || |   | |    ||   | ||||| ||  ||||||  ||||
Db   1162 GCCGGCCTGCTCGCCGAGGCGGGCACCACGTTCAGCACGCCTTCCGGCACCCCC 1108


RESULT   8
US-08-920-827-13/c
; Sequence 13, Application US/08920827
; Patent No. 5770375
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/920,827
;      FILING DATE:  29-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/362,577
;      FILING DATE:  27-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
```

```
;       TELEFAX:  312/474-0448
;       TELEX:  25-3856
;   INFORMATION FOR SEQ ID NO:  13:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  9515 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  Genomic DNA
;     ORIGINAL SOURCE:
;       ORGANISM:  Pseudomonas aeruginosa
;       STRAIN:  Clinical Isolate P2-2
US-08-920-827-13


  Query Match             3.2%;  Score 35;  DB 2;  Length 9515;
  Best Local Similarity   56.5%;  Pred. No. 1.3;
  Matches   65;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy    318 gactgcagagaagctgtggagcaacttggtgccagcgccaacctgtcagaagagcgtctg 377
          |||  | | | ||| |  || || |   | | |  | |    ||  ||  ||| | ||
Db   1222 GACCTCGGTGGAGCCGGTGAACACCAGTGCGTCCACCTCCGGGTGCAAGCCGAGGGCCTT 1163

Qy    378 gccgtcctgctggcgggcacacacaccctgctccagcaggctctccggctgcccc 432
          |||| |||||| || |   | |    ||   | ||||| ||  ||||||  ||||
Db   1162 GCCGGCCTGCTCGCCGAGGCCCGGCACCACGTTCAGCACGCCTTCCGGCACCCCC 1108


RESULT  9
US-08-921-177-13/c
; Sequence 13, Application US/08921177
; Patent No. 5798211
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/921,177
;      FILING DATE:  29-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/362,577
;      FILING DATE:  27-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9515 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Pseudomonas aeruginosa
;      STRAIN:  Clinical Isolate P2-2
US-08-921-177-13


  Query Match             3.2%;  Score 35;  DB 2;  Length 9515;
  Best Local Similarity   56.5%;  Pred. No. 1.3;
  Matches   65;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy    318 gactgcagagaagctgtggagcaacttggtgccagcgccaacctgtcagaagagcgtctg 377
          |||  | | | ||| |  || || |   | | |  | |    ||  ||  ||| | ||
Db   1222 GACCTCGGTGGAGCCGGTGAACACCAGTGCGTCCACCTCCGGGTGCAAGCCGAGGGCCTT 1163

Qy    378 gccgtcctgctggcgggcacacacaccctgctccagcaggctctccggctgcccc 432
          |||| |||||| || |   | |    ||   | ||||| ||  ||||||  ||||
Db   1162 GCCGGCCTGCTCGCCGAGGCCCGGCACCACGTTCAGCACGCCTTCCGGCACCCCC 1108


RESULT  10
US-08-362-577C-13/c
; Sequence 13, Application US/08362577C
; Patent No. 5807673
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/362,577C
;      FILING DATE:  27-MAR-1995
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9515 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Pseudomonas aeruginosa
;      STRAIN:  Clinical Isolate P2-2
US-08-362-577C-13


  Query Match             3.2%;  Score 35;  DB 2;  Length 9515;
  Best Local Similarity   56.5%;  Pred. No. 1.3;
  Matches   65;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy    318 gactgcagagaagctgtggagcaacttggtgccagcgccaacctgtcagaagagcgtctg 377
          |||  | | | ||| |  || || |   | | |  | |    ||  ||  ||| | ||
```

```
Db   1222 GACCTCGGTGGAGCCGGTGAACACCAGTGCGTCCACCTCCGGGTGCAAGCCGAGGGCCTT 1163

Qy    378 gccgtcctgctggcgggcacacacaccctgctccagcaggctctccggctgcccc 432
          |||| |||||| || |   | |    ||   | ||||| ||  ||||||  ||||
Db   1162 GCCGGCCTGCTCGCCGAGGCCCGGCACCACGTTCAGCACGCCTTCCGGCACCCCC 1108


RESULT  11
US-08-920-828-13/c
; Sequence 13, Application US/08920828
; Patent No. 5853998
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/920,828
;      FILING DATE:  29-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/362,577
;      FILING DATE:  27-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9515 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Pseudomonas aeruginosa
;      STRAIN:  Clinical Isolate P2-2
US-08-920-828-13


  Query Match             3.2%;  Score 35;  DB 3;  Length 9515;
  Best Local Similarity   56.5%;  Pred. No. 1.3;
  Matches   65;  Conservative   0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy    318 gactgcagagaagctgtggagcaacttggtgccagcgccaacctgtcagaagagcgtctg 377
          |||  | | | ||| |  || || |   | | |  |  |    ||  ||  ||| | ||
Db   1222 GACCTCGGTGGAGCCGGTGAACACCAGTGCGTCCACCTCCGGGTGCAAGCCGAGGGCCTT 1163

Qy    378 gccgtcctgctggcgggcacacacaccctgctccagcaggctctccggctgcccc 432
          |||| |||||| || |   | |    ||   | ||||| ||  ||||||  ||||
Db   1162 GCCGGCCTGCTCGCCGAGGCCCGGCACCACGTTCAGCACGCCTTCCGGCACCCCC 1108


RESULT  12
US-08-076-011-1/c
; Sequence 1, Application US/08076011
; Patent No. 5521069
;  GENERAL INFORMATION:
;    APPLICANT:  ONDA, Haruo
;    APPLICANT:  KIMURA, Chiharu
;    APPLICANT:  OHKUBO, Shoichi
;    TITLE OF INVENTION:  NOVEL DNA AND USE THEREOF
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DAVID G. CONLIN; DIKE, BRONSTEIN,
;      ADDRESSEE:  ROBERTS & CUSHMAN
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  US
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/076,011
;      FILING DATE:  11-JUN-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/047,246
;      FILING DATE:  13-APR-1993
;      NAME:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/741,676
;      FILING DATE:  07-AUG-1991
;      NAME:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  RESNICK, DAVID S
;      REGISTRATION NUMBER:  34235
;      REFERENCE/DOCKET NUMBER:  41155-CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)523-3400
;      TELEFAX:  (617)523-6440
;      TELEX:  200291 STRE UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  17041 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(7540..7650, 9814..9945, 10421..10519,
;      LOCATION:  11602..11787)
US-08-076-011-1


  Query Match             3.2%;  Score 34.8;  DB 1;  Length 17041;
  Best Local Similarity   60.6%;  Pred. No. 2;
  Matches   57;  Conservative   0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy    989 ccttggaagtctgaggctggagaatgggaagtgtaagctgggcctggctttcatagtgag 1048
          | |||| || ||||||  |||| || |   | | |||    ||||||    |||||||||
Db  14797 CTTTGGGAGGCTGAGGTGGGAGGATCGCTTGAGCAAGACCAGCCTGGGCAACATAGTGAG 14738

Qy   1049 gctcagtgtcgaattaaagaggtaaagcaactat 1082
            |   |   | ||   || |  ||||  ||| ||
Db  14737 ACCTGGCCACTAAAATAAAAAATAAAAGAACAAT 14704


RESULT  13
US-08-724-394A-20
; Sequence 20, Application US/08724394A
; Patent No. 5872237
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  Feder, John N.
;    APPLICANT:  Kronmal, Gregory S.
;    APPLICANT:  Lauer, Peter M.
;    APPLICANT:  Ruddy, David A.
;    APPLICANT:  Thomas, Winston
;    APPLICANT:  Tsuchihashi, Zenta
;    APPLICANT:  Wolff, Roger K.
;    TITLE OF INVENTION:  Megabase Transcript Map:  No. 5872237el
;    TITLE OF INVENTION:  Sequences and Antibodies Thereto
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TOWNSEND and TOWNSEND and CREW LLP
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/724,394A
;      FILING DATE:  01-OCT-1996
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fitts, Renee A.
;      REGISTRATION NUMBER:  35,136
;      REFERENCE/DOCKET NUMBER:  017957-000100
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-576-0200
;      TELEFAX:  415-576-0300
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  246240 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..246240
;      OTHER INFORMATION:  /note= "HLA-H.CONTIG"
US-08-724-394A-20


  Query Match             3.2%;  Score 34.8;  DB 3;  Length 246240;
  Best Local Similarity   57.3%;  Pred. No. 7;
  Matches   63;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy     990 cttggaagtctgaggctggagaatgggaagtgtaagctgggcctggctttcatagtgagg 1049
           || || || |||||||  ||||||     |  ||     ||    | || ||  | |||
Db 233301 CTCGGGAGGCTGAGGCAAGAGAATCACTTGAATATAGGAGGGGGAGGTTGCAGTGAGAGT 233360

Qy    1050 ctcagtgtcgaattaaagaggtaaagcaactattaaaaaaaaaaaaaaaa 1099
             | | || ||  |||     ||| |||  | ||| || |||||| |||
Db 233361 GAGACTCTCAAAAAAAAACAACAAAACAAAAAGTAACAAGAAAAAAGAAA 233410


RESULT  14
US-08-724-394A-21
; Sequence 21, Application US/08724394A
; Patent No. 5872237
;  GENERAL INFORMATION:
;    APPLICANT:  Feder, John N.
;    APPLICANT:  Kronmal, Gregory S.
;    APPLICANT:  Lauer, Peter M.
;    APPLICANT:  Ruddy, David A.
;    APPLICANT:  Thomas, Winston
;    APPLICANT:  Tsuchihashi, Zenta
;    APPLICANT:  Wolff, Roger K.
;    TITLE OF INVENTION:  Megabase Transcript Map:  No. 5872237el
;    TITLE OF INVENTION:  Sequences and Antibodies Thereto
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TOWNSEND and TOWNSEND and CREW LLP
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/724,394A
;      FILING DATE:  01-OCT-1996
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fitts, Renee A.
;      REGISTRATION NUMBER:  35,136
;      REFERENCE/DOCKET NUMBER:  017957-000100
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-576-0200
;      TELEFAX:  415-576-0300
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  246240 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..246240
;      OTHER INFORMATION:  /note= "HLA-H.CONTIG"
US-08-724-394A-21


  Query Match             3.2%;  Score 34.8;  DB 3;  Length 246240;
  Best Local Similarity   57.3%;  Pred. No. 7;
  Matches   63;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy     990 cttggaagtctgaggctggagaatgggaagtgtaagctgggcctggctttcatagtgagg 1049
           || || || |||||||  ||||||     |  ||     ||    | || ||  | |||
Db 233301 CTCGGGAGGCTGAGGCAAGAGAATCACTTGAATATAGGAGGGGGAGGTTGCAGTGAGAGT 233360

Qy    1050 ctcagtgtcgaattaaagaggtaaagcaactattaaaaaaaaaaaaaaaa 1099
             | | || ||  |||     ||| |||  | ||| || |||||| |||
Db 233361 GAGACTCTCAAAAAAAAACAACAAAACAAAAAGTAACAAGAAAAAAGAAA 233410


RESULT  15
US-08-724-394A-22
; Sequence 22, Application US/08724394A
; Patent No. 5872237
;  GENERAL INFORMATION:
;    APPLICANT:  Feder, John N.
;    APPLICANT:  Kronmal, Gregory S.
;    APPLICANT:  Lauer, Peter M.
;    APPLICANT:  Ruddy, David A.
;    APPLICANT:  Thomas, Winston
;    APPLICANT:  Tsuchihashi, Zenta
;    APPLICANT:  Wolff, Roger K.
;    TITLE OF INVENTION:  Megabase Transcript Map:  No. 5872237el
;    TITLE OF INVENTION:  Sequences and Antibodies Thereto
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TOWNSEND and TOWNSEND and CREW LLP
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
```

```
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/724,394A
;      FILING DATE:  01-OCT-1996
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fitts, Renee A.
;      REGISTRATION NUMBER:  35,136
;      REFERENCE/DOCKET NUMBER:  017957-000100
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-576-0200
;      TELEFAX:  415-576-0300
;  INFORMATION FOR SEQ ID NO:  22:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  246240 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..246240
;      OTHER INFORMATION:  /note= "HLA-H.CONTIG"
US-08-724-394A-22


  Query Match             3.2%;  Score 34.8;  DB 3;  Length 246240;
  Best Local Similarity   57.3%;  Pred. No. 7;
  Matches   63;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy    990 cttggaagtctgaggctggagaatgggaagtgtaagctgggcctggctttcatagtgagg 1049
          || || || |||||||  ||||||     |  ||     ||    | || ||  | |||
Db 233301 CTCGGGAGGCTGAGGCAAGAGAATCACTTGAATATAGGAGGGGGAGGTTGCAGTGAGAGT 233360

Qy   1050 ctcagtgtcgaattaaagaggtaaagcaactattaaaaaaaaaaaaaaa 1099
          | | || ||  |||     ||| |||  | ||| || |||||| |||
Db 233361 GAGACTCTCAAAAAAAAACAACAAAACAAAAAGTAACAAGAAAAAAGAAA 233410


Search completed: May 15, 2000, 01:50:08
Job time: 10796 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: May 14, 2000, 23:02:16 ; Search time 84.19 Seconds
(without alignments)
3268.931 Million cell updates/sec

Title: US-09-223-796-1
Perfect score: 1100
Sequence: 1 gcacgagccacagccagcta.....attaaaaaaaaaaaaaaaa 1100

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1100	100.0	1100	1	V18890	Rat Hypertension r
	2	149.2	13.6	701	1	V18891	Human Hypertension
	3	41.4	3.8	218	1	T32035	Proliferation-inhi
c	4	39.8	3.6	849	1	V06127	Viral infection ge
	5	37.4	3.4	6727	1	T88014	Murine IL-5 cDNA g
c	6	36.6	3.3	2858	1	Q75270	Unique 2.9 kb geno
c	7	36.6	3.3	5687	1	T65001	Mouse cell cycle r
c	8	36.4	3.3	4736	1	X23315	Mouse 1-alpha-OHas
	9	36.2	3.3	602	1	X20419	Human secreted pro
c	10	36.2	3.3	1371	1	V33755	Mouse neutrophil e
	11	36.2	3.3	2851	1	Q72975	Cancer repression
	12	36.2	3.3	5150	1	T78598	AS-30D tumour Type
c	13	36.2	3.3	14180	1	V05159	Mouse butyrophilin
	14	35.8	3.3	921	1	X51728	DNA encoding a hum
	15	35.4	3.2	1089	1	X37460	Human secreted pro
	16	35.4	3.2	2880	1	X18193	hNBC3a coding sequ
	17	35.4	3.2	3912	1	X18192	hNBC3 coding seque
	18	35	3.2	2458	1	T89399	Methods for diagno
	19	35	3.2	8353	1	T89396	Methods for diagno
c	20	35	3.2	9515	1	Q55145	Pseudomonas aerugi
c	21	34.8	3.2	17041	1	Q21065	Genomic DNA of hum
c	22	34.8	3.2	17041	1	Q67057	PACPAP38 DNA. PACA
c	23	34.8	3.2	237326	1	V57903	Hereditary haemoch
	24	34.2	3.1	2560	1	X18941	Human U62317 gene
	25	33.8	3.1	1573	1	V43607	Human secreted pro
	26	33.8	3.1	4630	1	V84139	Human prothrombina
	27	33.6	3.1	1698	1	X23318	Mouse 1-alpha-OHas
	28	33.6	3.1	3323	1	Q48675	Human myotonic dys
	29	33.6	3.1	4105	1	X23316	Mouse 1-alpha-OHas
	30	33.4	3.0	2344	1	T89885	Drosophila frizzle
	31	33	3.0	13206	1	Q05243	Rat acyl peptide h
	32	33	3.0	50000	1	X23517	Human kidney amino
c	33	32.8	3.0	392	1	Q61201	Human brain Expres
	34	32.8	3.0	6306	1	Q54841	Sequence encoding
	35	32.8	3.0	6579	1	T77783	cDNA encoding LexA
	36	32.8	3.0	6819	1	T77782	cDNA encoding GAL4
c	37	32.8	3.0	10897	1	T09187	MuTu putative onco
	38	32.6	3.0	655	1	X37545	Human secreted pro
c	39	32.6	3.0	1294	1	T60587	Protein cognate of
	40	32.6	3.0	3848	1	X27277	Prostate-tumour de
	41	32.6	3.0	5668	1	X27262	Prostate-tumour de
	42	32.6	3.0	5873	1	T61701	Human prostatic sp
c	43	32.4	2.9	3360	1	V19608	Mus musculus EPCR
	44	32.4	2.9	10169	1	Q53522	Cellulose synthase
	45	32	2.9	1908	1	V73002	Human adult brain

ALIGNMENTS

RESULT 1
V18890
ID V18890 standard; cDNA; 1100 BP.
AC V18890;
DT 09-JUN-1998 (first entry)
DE Rat Hypertension related calcium regulated gene (HCaRG) cDNA.
KW Hypertension related calcium regulated gene; HCaRG; rat parathyroid;
KW extracellular calcium concentration; antibody; hypertension; ss;
KW hyperthyroidism; osteoporosis; heart failure; diabetes; stroke;
KW cancer; inflammatory disease; asthma.
OS Rattus rattus.
FH Key Location/Qualifiers
FT CDS 131. .806
FT /*tag= a
FT /product= "Hypertension related calcium regulator"
FT misc_feature 173. .196
FT /*tag= b
FT /note= "EF-hand like motif"
PN WO9749807-A2.
PD 31-DEC-1997.
PF 23-JUN-1997; CA0439.
PR 21-JUN-1996; US-667495.
PA (GOSS/) GOSSARD F.
PA (HAME/) HAMET P.
PA (LEWA/) LEWANCZUK R.
PA (TREM/) TREMBLAY J.
PI Gossard F, Hamet P, Lewanczuk R, Tremblay J;
DR WPI; 98-077171/07.
DR P-PSDB; W37723.
PT Hypertension related calcium regulated gene - useful to develop
PT products to treat or detect, e.g. hypertension, stroke,
PT osteoporosis, heart failure, cancer, diabetes or asthma
PS Claim 2; Pages 25-26; 46pp; English.
CC This nucleic acid sequence encodes the hypertension related calcium
CC regulated gene (HCaRG), which was isolated from the rat parathyroid.
CC Its expression is regulated by extracellular calcium concentration.
CC An antibody against the protein, can be used to detect or modulate
CC (e.g. enhance or inhibit) abnormal calcium levels. They can
CC specifically be used to detect or treat, e.g. hypertension,
CC hyperthyroidism, osteoporosis, heart failure, diabetes, stroke, cancer,
CC inflammatory disease, and asthma.
SQ Sequence 1100 BP; 260 A; 279 C; 315 G; 246 T;

Query Match 100.0%; Score 1100; DB 1; Length 1100;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 gcacgagccacagccagctaccgcggctaggttcctccaggtgcagagggcggtaaaggc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GCACGAGCCACAGCCAGCTACCGCGGCTAGGTTCCTCCAGGTGCAGAGGGCGGTAAAGGC 60

Qy      61 ttggtttgtatttgtaatgcaactgtggttaggaccttctcttcggactggtcaagaaac 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TTGGTTTGTATTTGTAATGCAACTGTGGTTAGGACCTTCTCTTCGGACTGGTCAAGAAAC 120

Qy     121 gggaagaaaggatgtctgctttgggggctgcagctccatacttgcaccatcccgctgaca 180
```

```
Db    121 GGGAAGAAAGGATGTCTGCTTTGGGGGCTGCAGCTCCATACTTGCACCATCCCGCTGACA 180

Qy    181 gtcacagtggccgggtcagtttcctgggttcccagccctctccagaagtgacggccgtgg 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GTCACAGTGGCCGGGTCAGTTTCCTGGGTTCCCAGCCCTCTCCAGAAGTGACGGCCGTGG 240

Qy    241 ctcagctcttgaaggacttagacaggagcaccttcagaaagttgttgaaacttgtagtcg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CTCAGCTCTTGAAGGACTTAGACAGGAGCACCTTCAGAAAGTTGTTGAAACTTGTAGTCG 300

Qy    301 gggccctgcatgggaaagactgcagagaagctgtggagcaacttggtgccagcgccaacc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GGGCCCTGCATGGGAAAGACTGCAGAGAAGCTGTGGAGCAACTTGGTGCCAGCGCCAACC 360

Qy    361 tgtcagaagagcgtctggccgtcctgctggcgggcacacacacccctgctccagcaggctc 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGTCAGAAGAGCGTCTGGCCGTCCTGCTGGCGGGCACACACACCCTGCTCCAGCAGGCTC 420

Qy    421 tccggctgccccctgctagtctaaagccagatgccttccaggaagagctccaggaacttg 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TCCGGCTGCCCCCTGCTAGTCTAAAGCCAGATGCCTTCCAGGAAGAGCTCCAGGAACTTG 480

Qy    481 gcattcctcaggatctaattggagatttggccagtttggcatttgggagtcaacgccctc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GCATTCCTCAGGATCTAATTGGAGATTTGGCCAGTTTGGCATTTGGGAGTCAACGCCCTC 540

Qy    541 ttctcgactctgtagcccaacagcagggatcctcgctgcctcacgtgtcttacttccggt 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TTCTCGACTCTGTAGCCCAACAGCAGGGATCCTCGCTGCCTCACGTGTCTTACTTCCGGT 600

Qy    601 ggcgggtggacgtggccatctcaaccagcgctcagtcccgctccctgcaaccgagtgttc 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GGCGGGTGGACGTGGCCATCTCAACCAGCGCTCAGTCCCGCTCCCTGCAACCGAGTGTTC 660

Qy    661 tcatgcagctgaagctcacagatggatctgcacaccgcttcgaggtgcccatagccaaat 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 TCATGCAGCTGAAGCTCACAGATGGATCTGCACACCGCTTCGAGGTGCCCATAGCCAAAT 720

Qy    721 ttcaggagctgcggtacagtgtagccttggtccttaaggagatggcagaactggagaaga 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 TTCAGGAGCTGCGGTACAGTGTAGCCTTGGTCCTTAAGGAGATGGCAGAACTGGAGAAGA 780

Qy    781 agtgtgagcgcaaactgcaggactgactgaaccctggtactgtgggtgctgaagctggta 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 AGTGTGAGCGCAAACTGCAGGACTGACTGAACCCTGGTACTGTGGGTGCTGAAGCTGGTA 840

Qy    841 ccagaacacagccccccactggtgatgagcccaactccattgaggtcctgcatgtgagaa 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 CCAGAACACAGCCCCCCACTGGTGATGAGCCCAACTCCATTGAGGTCCTGCATGTGAGAA 900

Qy    901 cgtattttaagtgaaaagacagcgggactttcaggttttgttttatgagtcaacagctgg 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 CGTATTTTAAGTGAAAAGACAGCGGGACTTTCAGGTTTTGTTTTATGAGTCAACAGCTGG 960

Qy    961 gcagggtggcacagtttataatctcagcccttggaagtctgaggctggagaatgggaagt 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 GCAGGGTGGCACAGTTTATAATCTCAGCCCTTGGAAGTCTGAGGCTGGAGAATGGGAAGT 1020

Qy   1021 gtaagctgggcctggctttcatagtgaggctcagtgtcgaattaaagaggtaaagcaact 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 GTAAGCTGGGCCTGGCTTTCATAGTGAGGCTCAGTGTCGAATTAAAGAGGTAAAGCAACT 1080

Qy   1081 attaaaaaaaaaaaaaaaaa 1100
          ||||||||||||||||||||
Db   1081 ATTAAAAAAAAAAAAAAAAA 1100


RESULT   2
V18891
ID   V18891 standard; cDNA; 701 BP.
AC   V18891;
DT   09-JUN-1998  (first entry)
DE   Human Hypertension related calcium regulated gene (HCaRG) cDNA.
KW   Hypertension related calcium regulated gene; HCaRG; human parathyroid;
KW   extracellular calcium concentration; antibody; hypertension; ss;
KW   hyperthyroidism; osteoporosis; heart failure; diabetes; stroke;
KW   cancer; inflammatory disease; asthma.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   misc_feature    224. .245
FT                   /*tag= a
FT                   /note= "EF-hand like motif"
PN   WO9749807-A2.
PD   31-DEC-1997.
PF   23-JUN-1997; CA0439.
PR   21-JUN-1996; US-667495.
PA   (GOSS/) GOSSARD F.
PA   (HAME/) HAMET P.
PA   (LEWA/) LEWANCZUK R.
PA   (TREM/) TREMBLAY J.
PI   Gossard F, Hamet P, Lewanczuk R, Tremblay J;
DR   WPI; 98-077171/07.
PT   Hypertension related calcium regulated gene - useful to develop
PT   products to treat or detect, e.g. hypertension, stroke,
PT   osteoporosis, heart failure, cancer, diabetes or asthma
PS   Claim 3; Pages 27-28; 46pp; English.
CC   This nucleic acid sequence encodes the hypertension related calcium
CC   regulated gene (HCaRG), which was isolated from the human parathyroid.
CC   Its expression is regulated by extracellular calcium concentration.
CC   The sequence displays 75 percent homology to the rat form, with
CC   conserved areas such as the EF-hand like motifs and the initiating
CC   codon.  An antibody against the protein, can be used to detect or
CC   modulate (e.g. enhance or inhibit) abnormal calcium levels.  They can
CC   specifically be used to detect or treat, e.g. hypertension,
CC   hyperthyroidism, osteoporosis, heart failure, diabetes, stroke, cancer,
CC   inflammatory disease, and asthma.
SQ   Sequence   701 BP;    137 A;    204 C;    217 G;    139 T;


  Query Match           13.6%;  Score 149.2;  DB 1;  Length 701;
  Best Local Similarity   70.1%;  Pred. No. 8.9e-33;
  Matches  213;  Conservative    0;  Mismatches   90;  Indels    1;  Gaps    1;

Qy     88 gttaggaccttctcttcggactggtcaagaaacgggaagaaaggatgtctgctttggggg 147
          ||   ||||| ||   | | ||| |||| |    | || |   ||||||||||||||||
Db    138 GTCTCGACCTCCTGGGCTGGCTGATCAAAGAGGAAGCAGCAGCAATGTCTGCTTTGGGGA 197

Qy    148 ctgcagctccatacttgcaccatcccgctgacagtcacagtggccgggtcagtttcctgg 207
          |||||  |||||||| |||| ||||| | ||| ||||||||||||||| || ||||||| |||
Db    198 CTGCAACTCCATACCTGCATCATCCTGGTGATAGTCACAGTGGCCGAGTGAGTTTCTTGG 257

Qy    208 gttcccagccctctccagaagtgacggccgtggctcagctcttgaaggacttagacagga 267
          |  ||||||   ||||||| ||| | ||  |||| ||||| |   |||| |||||||||
Db    258 GGGCCCAGCTTCCTCCAGAGGTGGCAGCAATGGCCCAGCTACTAGGGGACCTAGACAGGA 317

Qy    268 gcaccttcagaaagttg-ttgaaacttgtagtcggggccctgcatgggaaagactgcaga 326
          |||| |||||||||||| |||||  |||| ||| |   |||||| ||| | || ||| ||
Db    318 GCACGTTCAGAAAGTTGCTTGAAGTTTGTGGTCAGCAGCCTGCAGGGGGAGGATTGCCGA 377

Qy    327 gaagctgtggagcaacttggtgccagcgccaacctgtcagaagagcgtctggccgtcctg 386
          ||   |          |||| | |||||||||||||| | || ||||   | |   |||
Db    378 GAGTGTTGCAGCGTNTTTGGGGTCAGCGCCAACCTGCCGGAGGAGCAGTTNGGTTGCCTT 437

Qy    387 ctgg 390
            | |
Db    438 GTTG 441


RESULT   3
T32035
ID   T32035 standard; DNA; 218 BP.
AC   T32035;
DT   13-OCT-1996  (first entry)
DE   Proliferation-inhibiting gene sequence.
KW   Cell proliferation; cancer; viral infection; AIDS; HIV;
```

KW mole; wart; psoriasis; chromosome 4; gene therapy; diagnosis; ss.
OS Homo sapiens.
PN WO9622688-A1.
PD 01-AUG-1996.
PF 26-JAN-1996; U01344.
PR 26-JAN-1995; US-378289.
PA (BAYU) BAYLOR COLLEGE MEDICINE.
PI Bertram MJ, Ning Y, Pereira-Smith OM;
DR WPI; 96-362381/36.
PT Isolated human chromosome 4 fragment - used for therapy and
PT diagnosis of cell proliferative disorders, e.g. cancer, AIDS, moles,
PT warts and psoriasis.
PS Claim 4; Page 31a; 40pp; English.
CC 3 Human chromosome-4 fragments (T32033-35) comprise genetic
CC determinants that are capable of inhibiting or arresting the
CC proliferative capacity of a cell. They were identified using a
CC microcell-mediated chromosome transfer protocol. The fragments
CC are located at 4q26-28 or 4q33-34.1. They can be used in gene
CC therapy of cancer and diseases, such as AIDS, that involve
CC proliferative cells. They can also be used to diagnose the extent
CC and/or severity of malignancy.
SQ Sequence 218 BP; 69 A; 31 C; 63 G; 55 T;

Query Match 3.8%; Score 41.4; DB 1; Length 218;
Best Local Similarity 58.5%; Pred. No. 0.014;
Matches 72; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy 977 tataatctcagcccttggaagtctgaggctggagaatgggaagtgtaagctgggcctggc 1036
Db 36 TTTAATCCCAGCAATTGGGAGGCAGAGGCAGGCATATTTTGAGTTTGAGGCCAGCCTGGT 95

Qy 1037 tttcatagtgaggctcagtgtcgaattaaagaggtaaagcaactattaaaaaaaaaaaaa 1096
Db 96 TTACAGAGTGAGTTTCAGGACAGTCAGGGATATACAGAGAAACCCTGTTTCAGACAAAAA 155

Qy 1097 aaa 1099
Db 156 AAA 158

RESULT 4
V06127/c
ID V06127 standard; DNA; 849 BP.
AC V06127;
DT 05-MAY-1998 (first entry)
DE Viral infection gene SEQ ID NO:47.
KW Viral infection; tumour suppressor; cellular gene; rat; cancer;
KW serum protein; inhibitor; malignant phenotype; HIV; influenza;
KW hepatitis; retrovirus; immunodeficiency; ds.
OS Rattus sp.
PN WO9739119-A1.
PD 23-OCT-1997.
PF 11-APR-1997; U06067.
PR 15-APR-1996; US-015334.
PA (UYVA-) UNIV VANDERBILT.
PI Dubois RN, Organ EL, Rubin DH;
DR WPI; 97-526456/48.
PT Genes involved in viral infection and tumour suppression - used to
PT develop products for reducing or preventing viral infection or for
PT suppressing tumours
PS Claim 1; Page 64-65; 101pp; English.
CC The present sequence represents a viral infection gene. The present
CC invention describes nucleic acid sequences isolated from rat. The
CC sequences of the invention comprise 70 viral infection (VI) genes and
CC 8 tumour suppressor (TS) genes. Propagating cell cultures in the
CC absence of the serum protein (SP) allows selective elimination of cells
CC persistently infected with a virus from the cell culture. Inhibitors of
CC the SP can be used for reducing or inhibiting a viral infection.
CC Inhibitors of the TS gene products can be used to suppress a malignant
CC phenotype (MP). The methods and inhibitors can be used with viruses
CC such as HIV, influenza, hepatitis virus or animal retroviruses such as
CC simian immunodeficiency virus, avian immunodeficiency virus, bovine
CC immunodeficiency virus, feline immunodeficiency virus, equine infectious
CC anemia virus, caprine arthritis encephalitis virus or visna virus.
CC Because the identified genes are non-essential to cell survival, the
CC treatment methods can be used in subjects without serious detrimental
CC effects to the subjects.
SQ Sequence 849 BP; 209 A; 169 C; 200 G; 216 T;

Query Match 3.6%; Score 39.8; DB 1; Length 849;
Best Local Similarity 57.1%; Pred. No. 0.07;
Matches 68; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy 938 ttgttttatgagtcaacagctgggcagggtggcacagtttataatctcagcccttggaag 997
Db 812 TGGGATTAGGAGGAAGTNNCTGGACATGGTTTACATTTGGGTCATCTTAACACCTAGGAG 753

Qy 998 tctgaggctggagaatgggaagtgtaagctgggcctggctttcatagtgaggctcagtg 1056
Db 752 GCTGAGACAGGAGGATCACAAGTTTAAGGCCTGACTGGGCTGCATNNTGAGATCCTGTG 694

RESULT 5
T88014
ID T88014 standard; DNA; 6727 BP.
AC T88014;
DT 22-APR-1998 (first entry)
DE Murine IL-5 cDNA genomic fusion gene.
KW Murine CD3delta; transcription control sequence; transgenic mouse; TCS;
KW T cell; interleukin-5; IL-5; thymocyte; disease model; therapeutic agent;
KW screening; eosinophil; chimeric; fusion gene; ss.
OS Chimeric - Mus sp.
PN WO9738086-A2.
PD 16-OCT-1997.
PF 09-APR-1997; U05932.
PR 09-APR-1996; US-629643.
PA (MAYO-) MAYO FOUNDATION.
PI Lee JJ, Lee NA;
DR WPI; 97-512706/47.
PT Transgenic mouse with interleukin-5 gene integrated into its genome
PT - under control of thymocyte and T cell, lung or basal keratinocyte
PT specific regulator, useful as disease model or to screen therapeutic
PT agent
PS Example 2; Page 73-75; 127pp; English.
CC This is a murine interleukin-5 (IL-5) cDNA genomic fusion gene. This
CC transgenic construct contain a chimeric pIL-5.EXPDNA sequence, comprising
CC thymocyte/T cell, lung or basal keratinocyte specific transcription
CC control sequence (TCS) and a segment encoding interleukin-5 (IL-5), which
CC lacks endogenous 5'-control sequences when the TCS is thymocyte/T cell
CC specific. The TCS is usually from a murine CD3delta gene. The chimeric
CC DNA sequence is integrated into the genome and causes IL-5 to be
CC expressed at a eosinophil related disease inducing level, specifically a
CC serum level at least 50 times greater than the wild type. The mice can be
CC used as models of IL-5 associated disease, e.g. asthma, eosinophilic
CC pneumonia, myalgia, Loffler's syndrome, allergy, emphysema, pulmonary
CC fibrosis, Wagener's granulomatosis, adult respiratory distress syndrome
CC (ARDS), bacterial or fungal infection, leukaemia, rheumatoid arthritis,
CC atopic dermatitis or contact hypersensitivity, or to screen for IL-5
CC antagonists which are potential therapeutic agents for an IL-5 associated
CC disease. These mice constitutively express IL-5 at high levels in
CC specific tissues.
SQ Sequence 6727 BP; 1765 A; 1538 C; 1414 G; 2010 T;

Query Match 3.4%; Score 37.4; DB 1; Length 6727;
Best Local Similarity 50.9%; Pred. No. 0.79;
Matches 89; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

Qy 921 agcgggactttcaggttttgttttatgagtcaacagctgggcagggtggcacagtttata 980
Db 1844 AGGTAGATATGAAAGTTTGTGCTTATAAATAAAAAGAATATGAGAGTGGCAAATTATGTA 1903

Qy 981 atctcagcccttggaagtctgaggctggagaatgggaagtgtaagctgggcctggctttc 1040

```
Db   1904 ATCCCAGTACTTGGGAGCCAAAGGCAGGGGTAGTCTGAGTCTAGGGCCAGCTTAGATACA 1963

Qy   1041 atagtgaggctcagtgtcgaattaaagaggtaaagcaactattaaaaaaaaaaaa 1095
          |   |  |   |  | | ||    ||||  ||  |  ||||| |  | |
Db   1964 TTGCCCTGTATGTATCAAAAGTAAATCCTATAAATAAATAAACAAAAACATTAGA 2018


RESULT   6
Q75270/c
ID   Q75270 standard; DNA; 2858 BP.
AC   Q75270;
DT   10-JUL-1995  (first entry)
DE   Unique 2.9 kb genomic DNA fragment from human metastatic tumour.
KW   metastasis; cancer; tumour cell; probe; diagnosis; identification;
KW   malignant tumour; ds.
OS   Homo sapiens.
PN   WO9428129-A.
PD   08-DEC-1994.
PF   27-MAY-1994; G01160.
PR   28-MAY-1993; GB-011130.
PA   (ISIS-) ISIS INNOVATION LTD.
PI   Tarin D;
DR   WPI; 95-022801/03.
PT   New human tumour metastasis gene - used to develop prods for
PT   diagnosis or assessing the prognosis of tumour metastasis in a
PT   patient.
PS   Claim 1; Page 14-15; 22pp ; English.
CC   This sequence encodes a protein (sequence not given in the
CC   specification) which is expressed in malignant human tumours and
CC   their metastases. The DNA sequence has the potential to be a
CC   valuable probe for the accurate assessment of the prognosis of
CC   patients with malignant tumours, by examination of a tiny biopsy
CC   sample, or even a few cells obtd. by fine needle aspiration, and
CC   thus to influence therapy. No homology to any known coding regions
CC   was found using EMBL and Genbank databases. Three possible exons
CC   were predicted, one in the forward strand in frame 2 (between bases
CC   536 and 942) and two in the reverse strand, in frames 1 (between bases
CC   2143 and 2398) and 2 (between bases 1625 and 1907). PCR primers
CC   were designed to be used with the RT-PCR technique to amplify homologous
CC   mRNA sequences from RNA extracted from metastatic human tumour
CC   cell lines (see Q75271 and Q79572-4).
SQ   Sequence   2858 BP;    708 A;    763 C;    665 G;    722 T;


  Query Match             3.3%;  Score 36.6;  DB 1;  Length 2858;
  Best Local Similarity   57.6%;  Pred. No. 0.93;
  Matches   87;  Conservative    0;  Mismatches   59;  Indels    5;  Gaps    1;

Qy    955 agctgggcagggtggcacagtttataatctcagcccttggaagtctgaggctggagaatg 1014
          ||| ||||  |||||| |   |  ||||| |||| || ||||| | ||||| ||   ||
Db   2075 AGCCGGGCGTGGTGGCCCTACTGTTAATCCCAGCACTCGGAAGGCAGAGGCAGGCAGATT 2016

Qy   1015 ggaagtgtaagctgggcctggctttcatagtgaggctcagtgtc-----gaattaaagag 1069
               | ||    ||||||  | || |||   | ||      |     |   || | ||
Db   2015 TCTGACTTTAGGCCAGCCTGGTCTACAAAGTTGAGTTCCAGAACAGCCAGGGCTACACAG 1956

Qy   1070 gtaaagcaactattaaaaaaaaaaaaaaaa 1100
          |||  |   | | |||||||  ||||||||
Db   1955 AGAAACCCTGTCTCAAAAAAACCAAAAAAAA 1925


RESULT   7
T65001/c
ID   T65001 standard; DNA; 5687 BP.
AC   T65001;
DT   28-MAY-1997  (first entry)
DE   Mouse cell cycle regulatory protein SPA-1 genomic DNA (exons 1-4).
KW   Cell division cycle; Span-N; Span-C; lymphocyte differentiation;
KW   mouse; murine; ds.
OS   Mus sp.
FH   Key             Location/Qualifiers
FT   exon            3109. .3284
FT                   /*tag=  a
FT                   /number= 1
FT   exon            3764. .4555
FT                   /*tag=  b
FT                   /number= 2
FT   exon            5147. .5273
FT                   /*tag=  c
FT                   /number= 3
FT   exon            5383. .5524
FT                   /*tag=  d
FT                   /number= 4
PN   J08217797-A.
PD   27-AUG-1996.
PF   17-FEB-1995; 051999.
PR   30-MAY-1994; JP-139513.
PR   20-OCT-1994; JP-279712.
PR   13-DEC-1994; JP-332520.
PA   (MINA/) MINATO N.
DR   WPI; 96-439574/44.
PT   Gene encoding cell division-regulatory protein, SPA-1 - useful as
PT   regulator of cell division and differentiation of lymphocytes
PS   Claim 13; Page 12-14; 26pp; Japanese.
CC   The present partial genomic DNA sequence comprises exons 1-4 for a
CC   gene encoding cell division-regulating protein (designated SPA-1),
CC   which is not expressed in the resting phase but is expressed after
CC   the growth phase in the cell cycle of mammalian cells. SPA-1 is
CC   useful as a regulator of cell division or for differentiation
CC   of lymphocytes.
SQ   Sequence   5687 BP;    1201 A;    1574 C;    1632 G;    1280 T;


  Query Match             3.3%;  Score 36.6;  DB 1;  Length 5687;
  Best Local Similarity   56.1%;  Pred. No. 1.2;
  Matches   69;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy    977 tataatctcagcccttggaagtctgaggctggagaatgggaagtgtaagctgggcctggc 1036
          | ||||| ||| || || | || ||||| ||   ||    ||| |||    ||||||
Db   2712 TTTAATCCTAGCACTGGGGAATCAGAGGCAGGCAGATTTCGAGTTCAAGGCCAGCCTGGT 2653

Qy   1037 tttcatagtgaggctcagtgtcgaattaaagaggtaaagcaactattaaaaaaaaaaaaa 1096
          | || ||||||   |||  | |         |   | || |||  |      ||||| ||
Db   2652 CTACAGAGTGAGTTCCAGGATAGCCAGGGCTACACAGAGAAACCCTGTCTTGAAAAACAA 2593

Qy   1097 aaa 1099
          |||
Db   2592 AAA 2590


RESULT   8
X23315/c
ID   X23315 standard; DNA; 4736 BP.
AC   X23315;
DT   11-JUN-1999  (first entry)
DE   Mouse 1-alpha-OHase promoter region DNA #1.
KW   1-alpha-OHase; vitamin D 1-alpha-hydroxylase; vitamin D pathway;
KW   renal cytochrome P450 enzyme; 25-hydroxyvitamin D; human; mouse; rat;
KW   diagnosis; treatment; disorder; endocrine; promoter; ss.
OS   Mus sp.
PN   WO9907835-A2.
PD   18-FEB-1999.
PF   06-AUG-1998; CA0758.
PR   06-AUG-1997; US-906791.
PA   (SHRI-) SHRINERS HOSPITALS FOR CHILDREN.
PI   Glorieux FH, St-Arnaud R;
DR   WPI; 99-190048/16.
PT   New polynucleotides encoding 25-hydroxyvitamin D 1-^a-hydroxylase -
PT   useful to detect and treat vitamin D-related disorders
PS   Example 4; Page 77-79; 85pp; English.
CC   This invention describes novel 25-hydroxyvitamin D 1-alpha-hydroxylase
CC   (1-alpha-OHase), a renal cytochrome P450 enzyme of the vitamin D
CC   pathway, and polynucleotides encoding the enzyme from human, mouse and
CC   rat tissue. Molecules of the invention are used to diagnose and treat
CC   Vitamin D-related disorders, to produce vitamin D metabolites, and to
CC   identify modulators of the enzyme expression. The invention describes
```

CC an animal cell transformed with the enzyme promoter which is used to
CC identify compounds which modulate activity of the polynucleotide
CC encoding the enzyme, and to modulate the production of vitamin D in
CC patients with vitamin D endocrine disorders In the vitamin D
CC hydroxylation pathway, the enzyme catalyses conversion of
CC 25-hydroxyvitamin D to 1 alpha 25(OH)2D3.
SQ Sequence 4736 BP; 1061 A; 1375 C; 1160 G; 1140 T;

Query Match 3.3%; Score 36.4; DB 1; Length 4736;
Best Local Similarity 56.8%; Pred. No. 1.3;
Matches 67; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy 953 acagctgggcagggtggcacagtttataatctcagcccttggaagtctgaggctggagaa 1012
Db 4479 ACAGCTGGAAGTGGTAGCTCAAGTCTGTAACCCAATACTTGGGAAGGTGAGGCAAGGAGG 4420

Qy 1013 tgggaagtgtaagctgggcctggctttcatagtgaggctcagtgtcgaattaaagagg 1070
Db 4419 TTGTGAGTTCAAGGACAGCATGGAATACATAGCAAGACTCTGTCTCAAAAAGAAAAAG 4362

RESULT 9
X20419
ID X20419 standard; DNA; 602 BP.
AC X20419;
DT 04-MAY-1999 (first entry)
DE Human secreted protein gene 8.
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS Homo sapiens.
PN WO9906423-A1.
PD 11-FEB-1999.
PF 29-JUL-1998; U15949.
PR 19-AUG-1997; US-056730.
PR 30-JUL-1997; US-054209.
PR 30-JUL-1997; US-054211.
PR 30-JUL-1997; US-054212.
PR 30-JUL-1997; US-054213.
PR 30-JUL-1997; US-054214.
PR 30-JUL-1997; US-054215.
PR 30-JUL-1997; US-054217.
PR 30-JUL-1997; US-054218.
PR 30-JUL-1997; US-054234.
PR 30-JUL-1997; US-054236.
PR 18-AUG-1997; US-055968.
PR 18-AUG-1997; US-055969.
PR 18-AUG-1997; US-055972.
PR 19-AUG-1997; US-056534.
PR 19-AUG-1997; US-056543.
PR 19-AUG-1997; US-056554.
PR 19-AUG-1997; US-056561.
PR 19-AUG-1997; US-056727.
PR 19-AUG-1997; US-056729.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Carter KC, Endress GA, Fan P, Feng P, Kyaw H, Lafleur DW,
PI Li Y, Moore PA, Rosen CA, Ruben SM, Shi Y, Wei Y,
PI Zeng Z;
DR WPI; 99-153691/13.
DR P-PSDB; Y00265.
PT New isolated human genes and the secreted polypeptides they encode -
PT useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
PS Claim 1; Page 214; 312pp; English.
CC This sequence represents a nucleic acid molecule which encodes a secreted
CC human protein. The gene number is given in the descriptor line. The gene
CC can be used to generate fusion proteins by linking to the gene to a human
CC immunoglobulin Fc portion (e.g. X20403) for increasing the stability of
CC the fused protein as compared to the human protein only.
CC The invention relates to 83 novel genes and their fragments (nucleic acid
CC sequences: X20412-X20499; amino acid sequences Y00258-Y00377) which
CC are useful for preventing, treating or ameliorating medical conditions
CC e.g. by protein or gene therapy. Also, pathological conditions can be
CC diagnosed by determining the amount of the new polypeptides in a sample
CC or by determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the 86 polynucleotides, based on
CC which tissues they are most highly expressed in (see X20412 for described
CC uses).
SQ Sequence 602 BP; 194 A; 133 C; 152 G; 123 T;

Query Match 3.3%; Score 36.2; DB 1; Length 602;
Best Local Similarity 62.9%; Pred. No. 0.63;
Matches 56; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

Qy 1012 atgggaagtgtaagctgggcctggctttcatagtgaggctcagtgtcgaattaaagaggt 1071
Db 481 ATGGCGACACTGACTCCAGCCTGGGTGCCAGAGCGAGACTCTTTGTCTCAAAAAAAAAAA 540

Qy 1072 aaagcaactattaaaaaaaaaaaaaaaaa 1100
Db 541 AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 569

RESULT 10
V33755/c
ID V33755 standard; DNA; 1371 BP.
AC V33755;
DT 02-FEB-1999 (first entry)
DE Mouse neutrophil elastase gene promoter.
KW Neutrophil elastase; mouse; transgenic animal; knockout animal;
KW emphysema; cystic fibrosis; infection; animal model; promoter; ds.
OS Mus musculus strain 129/SvJ.
PN WO9844091-A1.
PD 08-OCT-1998.
PF 30-MAR-1998; U06266.
PR 27-MAR-1998; US-049279.
PR 31-MAR-1997; US-042871.
PA (BARN-) BARNES-JEWISH HOSPITAL.
PI Belaaouaj A, Shapiro SD;
DR WPI; 98-557103/47.
PT New murine neutrophil elastase knockout transgenic mice - useful as
PT models for studying disorders such as pulmonary emphysema
PS Claim 11; Page 47-48; 77pp; English.
CC This is the nucleotide sequence of the promoter region of the
CC murine neutrophil elastase (NE) gene (see also V33754). NE is an
CC enzyme that has a specific role in host defence against
CC Gram-positive, but not Gram-negative, bacteria. Mouse NE genomic
CC DNA was isolated from a murine 129/SvJ genomic library using a
CC human NE probe. Vectors comprising NE nucleotide sequences,
CC portions of these sequences, and transformed host cells harboring
CC such vectors are also provided, as well as transgenic and knockout
CC mice. Transgenic animals deficient in expression of NE can be used
CC to study the physiological and pathological functions of NE and its
CC interaction with other enzymes having a potential relationship to
CC cardiopulmonary function and pathology, such as cystic fibrosis
CC transmembrane conductance regulator. They can also be used as
CC models for disorders such as pulmonary emphysema and for the role
CC of NE in combating infection. The promoter of the NE gene can be
CC used for studying the regulation of the gene, or in studies where
CC targetting of gene expression to neutrophils is desired.
SQ Sequence 1371 BP; 352 A; 303 C; 339 G; 377 T;

Query Match 3.3%; Score 36.2; DB 1; Length 1371;
Best Local Similarity 60.8%; Pred. No. 0.89;
Matches 76; Conservative 0; Mismatches 48; Indels 1; Gaps 1;

Qy 977 tataatctcagcccttggaagtctgaggctgg-agaatgggaagtgtaagctgggcctgg 1035

```
Db    934 TTTAACCCCAGCACTCGGGAGGCAGAGGCAGGCAGATTTCTGAGTTCGAGGCCAGCCTGG 875

Qy   1036 ctttcatagtgaggctcagtgtcgaattaaagaggtaaagcaactattaaaaaaaaaaaa 1095
          | || ||||||   |||         || | ||  ||| |   | || ||||| |||||
Db    874 TCTACAAAGTGAGTTCCAGGCCAAGGCTATACAGAGAAACCCTGTCTTGAAAAACAAAAA 815

Qy   1096 aaaaa 1100
          ||||
Db    814 CAAAA 810


RESULT  11
Q72975
ID   Q72975 standard; cDNA; 2851 BP.
AC   Q72975;
DT   15-JUN-1995  (first entry)
DE   Cancer repression gene.
KW   Cancer; anticancer agent; tumour repression; ss.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   cds             10. .411
FT                   /*tag= a
FT                   /product= Anticancer polypeptide.
PN   J06253848-A.
PD   13-SEP-1994.
PF   03-MAR-1993; 069231.
PR   03-MAR-1993; JP-069231.
PA   (AGEN ) AGENCY OF IND SCI & TECHNOLOGY.
PA   (ROHT-) ROHTO SEIYAKU KK.
DR   WPI; 94-328995/41.
DR   P-PSDB; R62634.
PT   Novel DNA encoding a peptide - useful as an anticancer, tumour
PT   repressing agent
PS   Claim 1; Page 5-6; 8pp; Japanese.
CC   The polypeptide encoded by this sequence (and derivatives having
CC   amino acid substitutions, additions or deletions) may be used as an
CC   anticancer/tumour repressing agent.
SQ   Sequence   2851 BP;    745 A;    674 C;    660 G;    769 T;


  Query Match             3.3%;  Score 36.2;  DB 1;  Length 2851;
  Best Local Similarity   59.0%;  Pred. No. 1.2;
  Matches   62;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy    996 agtctgaggctggagaatgggaagtgtaagctgggcctggctttcatagtgaggctcagt 1055
          || || ||| ||| ||  |||| | |    |||| |    ||||   || |||||
Db   2743 AGGCTAAGGTGGGATGATCATAAGTTTGAAGCTAGCCTAGGGCACATAACAAGTCTCAGA 2802

Qy   1056 gtcgaattaaagaggtaaagcaactattaaaaaaaaaaaaaaaa 1100
          | | ||    ||    |||||||   ||   | ||||||||||||
Db   2803 GGCCAACCTGAGCTACAAAGCAAGCTTTGTCATAAAAAAAAAAAA 2847


RESULT  12
T78598
ID   T78598 standard; DNA; 5150 BP.
AC   T78598;
DT   30-OCT-1997  (first entry)
DE   AS-30D tumour Type II hexokinase 4.3kbp proximal promoter region.
KW   Response element; Z-DNA; neoplasia; hexokinase II; glycolysis;
KW   cancer; gene therapy; diabetes; tumour; rat; ds.
OS   Rattus rattus.
FH   Key             Location/Qualifiers
FT   protein_bind    58. .63
FT                   /*tag= a
FT                   /label= Myb
FT                   /note= "Transcription factor binding site"
FT   protein_bind    118. .126
FT                   /*tag= b
FT                   /label= p53
FT                   /note= "Transcription factor binding site"
FT   protein_bind    171. .180
FT                   /*tag= c
FT                   /label= p53
FT                   /note= "Transcription factor binding site"
FT   protein_bind    208. .220
FT                   /*tag= d
FT                   /label= HNF-5,AP-3,c/ebp
FT                   /note= "Transcription factor binding site"
FT   protein_bind    231. .240
FT                   /*tag= e
FT                   /label= IRE
FT                   /note= "Transcription factor binding site"
FT   protein_bind    322. .327
FT                   /*tag= f
FT                   /label= Myb
FT                   /note= "Transcription factor binding site"
FT   protein_bind    453. .459
FT                   /*tag= g
FT                   /label= HNF-5
FT                   /note= "Transcription factor binding site"
FT   protein_bind    504. .523
FT                   /*tag= h
FT                   /label= Pea-3,SRE,Ap-2
FT                   /note= "Transcription factor binding site"
FT   repeat_region   529. .559
FT                   /*tag= i
FT                   /rpt_type= DIRECT
FT                   /note= "Motif with the potential to form Z-DNA
FT                   structure"
FT   protein_bind    605. .610
FT                   /*tag= j
FT                   /label= G1RE,Myc
FT                   /note= "Transcription factor binding site"
FT   protein_bind    640. .647
FT                   /*tag= k
FT                   /label= c/ebp
FT                   /note= "Transcription factor binding site"
FT   protein_bind    723. .728
FT                   /*tag= l
FT                   /label= Pea-3
FT                   /note= "Transcription factor binding site"
FT   protein_bind    741. .746
FT                   /*tag= m
FT                   /label= Pea-3
FT                   /note= "Transcription factor binding site"
FT   protein_bind    899. .906
FT                   /*tag= n
FT                   /label= Ap-1
FT                   /note= "Transcription factor binding site"
FT   protein_bind    1037. .1043
FT                   /*tag= o
FT                   /label= HNF-5
FT                   /note= "Transcription factor binding site"
FT   protein_bind    1076. .1084
FT                   /*tag= p
FT                   /label= Sp-1
FT                   /note= "Transcription factor binding site"
FT   protein_bind    1112. .1117
FT                   /*tag= q
FT                   /label= Pea-3
FT                   /note= "Transcription factor binding site"
FT   protein_bind    1414. .1419
FT                   /*tag= r
FT                   /label= PPAR
FT                   /note= "Transcription factor binding site"
FT   protein_bind    1634. .1640
FT                   /*tag= s
FT                   /label= Ap-1
FT                   /note= "Transcription factor binding site"
FT   protein_bind    1816. .1824
FT                   /*tag= t
FT                   /label= c/ebp,NF-IL6
FT                   /note= "Transcription factor binding site"
FT   protein_bind    2045. .2051
FT                   /*tag= u
```

```
FT                   /label= Ap-1
FT                   /note= "Transcription factor binding site"
FT   repeat_unit     2090. .2099
FT                   /*tag=  v
FT                   /rpt_type= DIRECT
FT                   /label= I
FT   repeat_unit     2133. .2142
FT                   /*tag=  w
FT                   /rpt_type= DIRECT
FT                   /label= I
FT   protein_bind    2144. .2152
FT                   /*tag=  x
FT                   /label= Sp-1
FT                   /note= "Transcription factor binding site"
FT   protein_bind    2165. .2171
FT                   /*tag=  y
FT                   /label= HNF-5
FT                   /note= "Transcription factor binding site"
FT   protein_bind    2325. .2332
FT                   /*tag=  z
FT                   /label= Ap-2
FT                   /note= "Transcription factor binding site"
FT   protein_bind    2402. .2409
FT                   /*tag=  aa
FT                   /label= Ap-2
FT                   /note= "Transcription factor binding site"
FT   protein_bind    2413. .2420
FT                   /*tag=  ab
FT                   /label= Ap-1
FT                   /note= "Transcription factor binding site"
FT   repeat_region   2593. .2636
FT                   /*tag=  ac
FT                   /rpt_type= DIRECT
FT   protein_bind    2757. .2766
FT                   /*tag=  ad
FT                   /label= p53
FT                   /note= "Transcription factor binding site"
FT   protein_bind    2776. .2784
FT                   /*tag=  ae
FT                   /label= NF-IL6,Ap-1
FT                   /note= "Transcription factor binding site"
FT   protein_bind    2866. .2873
FT                   /*tag=  af
FT                   /label= Ap-2
FT                   /note= "Transcription factor binding site"
FT   protein_bind    2908. .2919
FT                   /*tag=  ag
FT                   /label= Myb,Mbf-1,Mep-1,Mtf-1
FT                   /note= "Transcription factor binding site"
FT   protein_bind    2926. .2933
FT                   /*tag=  ah
FT                   /label= c/ebp
FT                   /note= "Transcription factor binding site"
FT   protein_bind    2951. .2956
FT                   /*tag=  ai
FT                   /label= Pea-3
FT                   /note= "Transcription factor binding site"
FT   protein_bind    2988. .2993
FT                   /*tag=  aj
FT                   /label= Myb
FT                   /note= "Transcription factor binding site"
FT   protein_bind    2998. .3003
FT                   /*tag=  ak
FT                   /label= Pea-3
FT                   /note= "Transcription factor binding site"
FT   protein_bind    3076. .3086
FT                   /*tag=  al
FT                   /label= SRE
FT                   /note= "Transcription factor binding site"
FT   protein_bind    3102. .3109
FT                   /*tag=  am
FT                   /label= Ap-2
FT                   /note= "Transcription factor binding site"
```

```
FT   protein_bind    3162. .3167
FT                   /*tag=  an
FT                   /label= Myb
FT                   /note= "Transcription factor binding site"
FT   protein_bind    3223. .3228
FT                   /*tag=  ao
FT                   /label= SRE
FT                   /note= "Transcription factor binding site"
FT   protein_bind    3256. .3267
FT                   /*tag=  ap
FT                   /label= Sp-1,Ap-2,SRE
FT                   /note= "Transcription factor binding site"
FT   protein_bind    3306. .3313
FT                   /*tag=  aq
FT                   /label= c/ebp
FT                   /note= "Transcription factor binding site"
FT   protein_bind    3352. .3360
FT                   /*tag=  ar
FT                   /label= NF-IL6
FT                   /note= "Transcription factor binding site"
FT   repeat_unit     3466. .3483
FT                   /*tag=  as
FT                   /rpt_type= DIRECT
FT                   /label= II
FT   repeat_unit     3492. .3509
FT                   /*tag=  at
FT                   /rpt_type= DIRECT
FT                   /label= III
FT   protein_bind    3507. .3513
FT                   /*tag=  au
FT                   /label= Ap-1
FT                   /note= "Transcription factor binding site"
FT   protein_bind    3702. .3709
FT                   /*tag=  av
FT                   /label= Ap-2
FT                   /note= "Transcription factor binding site"
```

```
  Query Match              3.3%;  Score 36.2;  DB 1;  Length 5150;
  Best Local Similarity   60.8%;  Pred. No. 1.5;
  Matches   59;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy     952 aacagctgggcagggtggcacagtttataatctcagcccttggaagtctgaggctggaga 1011
           || |   ||  ||| | ||  |  | ||| ||||||  ||  | |    ||||| ||||
Db    1071 AAAAATAGGCGAGGTAGACAGTGCCTGTAACCTCAGCGCTGAGGAAGTGGAGGCCGGAGG 1130

Qy    1012 atgggaagtgtaagctgggcctggctttcatagtgag 1048
           ||||||||   |||   | ||| | || |||  | ||
Db    1131 ATGGGAAGCTCAAGACTGTCCTTGGTTGCATGTTTAG 1167
```

```
RESULT  13
V05159/c
ID   V05159 standard; DNA; 14180 BP.
AC   V05159;
DT   20-JUL-1998  (first entry)
DE   Mouse butyrophilin gene.
KW   Butyrophilin; btn gene; promoter; mouse; transgenic animal;
KW   mammary gland-specific gene expression; breast cancer; diagnosis;
KW   ds.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   promoter        1. .4693
FT                   /*tag=  a
FT                   /note= "Claim 4"
FT   TATA_signal     4645. .4651
FT                   /*tag=  b
FT   prim_transcript 4611
FT                   /*tag=  c
FT                   /evidence= EXPERIMENTAL
FT                   /standard_name= transcription start site
FT                   /note= "most frequently used site"
FT   misc_signal     4645. .4651
FT                   /*tag=  d
```

```
FT                   /note= "AT-rich, TATAA-like element"
FT   misc_signal     4583. .4588
FT                   /*tag= e
FT                   /note= "AT-rich, TATAA-like element"
FT   misc_signal     4597. .4603
FT                   /*tag= f
FT                   /note= "CCAAT-like element"
FT   misc_signal     4546. .4551
FT                   /*tag= g
FT                   /note= "CCAAT-like element"
FT   misc_signal     4533. .4537
FT                   /*tag= h
FT                   /note= "CCAAT-like element"
FT   prim_transcript 4675
FT                   /*tag= i
FT                   /evidence= EXPERIMENTAL
FT                   /standard_name= transcription start site
FT   prim_transcript 4694
FT                   /*tag= j
FT                   /evidence= EXPERIMENTAL
FT                   /standard_name= transcription start site
FT   polyA_signal    13091. .13096
FT                   /*tag= k
FT                   /evidence= EXPERIMENTAL
FT   prim_transcript 13097. .13199
FT                   /*tag= l
FT                   /evidence= EXPERIMENTAL
FT                   /standard_name= 3' end of transcript
FT   CDS             4923. .11397
FT                   /*tag= m
FT                   /evidence= EXPERIMENTAL
FT                   /note= "contains introns"
FT   sig_peptide     4923. .5001
FT                   /*tag= n
FT   CDS             4650. .4651
FT                   /*tag= o
FT                   /evidence= EXPERIMENTAL
FT                   /standard_name= translational initiation codon
FT   CDS             4743. .4745
FT                   /*tag= p
FT                   /evidence= EXPERIMENTAL
FT                   /standard_name= translational initiation codon
FT   CDS             4765. .4767
FT                   /*tag= q
FT                   /evidence= EXPERIMENTAL
FT                   /standard_name= translational initiation codon
FT   CDS             4776. .4778
FT                   /*tag= r
FT                   /evidence= EXPERIMENTAL
FT                   /standard_name= translational initiation codon
FT   exon            4611. .5001
FT                   /*tag= s
FT                   /codon_start= 4923. .4925
FT                   /number= 1
FT   intron          5002. .5520
FT                   /*tag= t
FT                   /evidence= EXPERIMENTAL
FT                   /number= A
FT   exon            5521. .5871
FT                   /*tag= u
FT                   /number= 2
FT   intron          5872. .8332
FT                   /*tag= v
FT                   /evidence= EXPERIMENTAL
FT                   /number= B
FT   exon            8333. .8614
FT                   /*tag= w
FT                   /number= 3
FT   intron          8615. .9485
FT                   /*tag= x
FT                   /evidence= EXPERIMENTAL
FT                   /number= C
FT   exon            9486. .9635
FT                   /*tag= y
FT                   /number= 4
FT   intron          9636. .10206
FT                   /*tag= z
FT                   /evidence= EXPERIMENTAL
FT                   /number= D
FT   exon            10207. .10227
FT                   /*tag= aa
FT                   /number= 5
FT   intron          10228. .10320
FT                   /*tag= ab
FT                   /evidence= EXPERIMENTAL
FT                   /number= E
FT   exon            10321. .10347
FT                   /*tag= ac
FT                   /number= 6
FT   intron          10348. .10738
FT                   /*tag= ad
FT                   /evidence= EXPERIMENTAL
FT                   /number= F
FT   exon            10738. .13199
FT                   /*tag= ae
FT                   /number= 7
FT   repeat_unit     4807. .4814
FT                   /*tag= af
FT                   /rpt_type= INVERTED
FT   repeat_unit     12556. .12563
FT                   /*tag= ag
FT                   /rpt_type= INVERTED
FT   protein_bind    2948. .2960
FT                   /*tag= ah
FT                   /note= "NF1"
FT   protein_bind    complement (2971. .2979)
FT                   /*tag= ai
FT                   /note= "gamma-IRE"
FT   protein_bind    complement (3010. .3015)
FT                   /*tag= aj
FT                   /note= "myb"
FT   protein_bind    complement (3030. .3036)
FT                   /*tag= ak
FT                   /note= "YY1"
FT   protein_bind    complement (3115. .3123)
FT                   /*tag= al
FT                   /note= "gamma-IRE"
FT   protein_bind    3129. .3135
FT                   /*tag= am
FT                   /note= "YY1"
FT   protein_bind    3132. .3146
FT                   /*tag= an
FT                   /note= "NF1"
FT   protein_bind    3138. .3146
FT                   /*tag= ao
FT                   /note= "STAT"
FT   protein_bind    complement (3145. .3150)
FT                   /*tag= ap
FT                   /note= "GMCSF"
FT   protein_bind    3181. .3189
FT                   /*tag= aq
FT                   /note= "C/EBP"
FT   protein_bind    complement (3187. .3197)
FT                   /*tag= ar
FT                   /note= "C/EBP"
FT   protein_bind    3190. .3195
FT                   /*tag= as
FT                   /note= "PU.1"
FT   protein_bind    complement (3194. .3200)
FT                   /*tag= at
FT                   /note= "alpha-IFN"
FT   protein_bind    3195. .3203
FT                   /*tag= au
FT                   /note= "GMCSF"
FT   protein_bind    3253. .3259
FT                   /*tag= av
```

```
FT                   /note= "gamma-IRE"
FT   protein_bind    3370. .3378
FT                   /*tag=  aw
FT                   /note= "GRE"
FT   protein_bind    complement (3397. .3406)
FT                   /*tag=  ax
FT                   /note= "GRE"
FT   protein_bind    complement (3399. .3403)
FT                   /*tag=  ay
FT                   /note= "gamma-IRE"
FT   protein_bind    3419. .3427
FT                   /*tag=  az
FT                   /note= "STAT"
FT   protein_bind    3439. .3445
FT                   /*tag=  ba
FT                   /note= "gamma-IRE"
FT   protein_bind    3450. .3454
FT                   /*tag=  bb
FT                   /note= "GMCSF"
FT   protein_bind    3456. .3462
FT                   /*tag=  bc
FT                   /note= "STAT"
FT   protein_bind    3457. .3463
FT                   /*tag=  bd
FT                   /note= "C/EBP"
FT   protein_bind    complement (3464. .3468)
FT                   /*tag=  be
FT                   /note= "GMCSF"
FT   protein_bind    3465. .3470
FT                   /*tag=  bf
FT                   /note= "alpha-INF"
FT   protein_bind    complement (3499. .3506)
FT                   /*tag=  bg
```

Query Match 3.3%; Score 36.2; DB 1; Length 14180;
Best Local Similarity 53.1%; Pred. No. 2.3;
Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

```
Qy     956 gctgggcagggtggcacagtttataatctcagcccttggaagtctgaggctggagaatgg 1015
           | | ||   |||||    |  |||  |  |||| ||||||||   ||||| |||| ||
Db   12740 GATAGGGGTGGTGGATGTGGCTATGTTTCCAGCACTTGGAAGGTAGAGGCAGGAGGATCA 12681

Qy    1016 gaagtgtaagctgggcctggctttcatagtgaggctcagtgtcgaattaaagaggtaaag 1075
           | | |  ||     ||||    | |||||| |  |||  | |  ||  | | |   | |
Db   12680 GGAATTCAAAGCCAGCCTAAGCTACATAGTAAACCTCTATCTTAAAACACACACACACAC 12621

Qy    1076 caactattaaaaaaaaaaaaaaaa 1100
             ||    | | | | | | | | |
Db   12620 ACACACACACACACACACACACACA 12596
```

RESULT 14
X51728
ID X51728 standard; DNA; 921 BP.
AC X51728;
DT 17-JUN-1999 (first entry)
DE DNA encoding a human secreted protein.
KW Human secreted protein; cancer; immune disorder; infection;
KW inflammatory disorder; skin disorder; tumour; atherosclerosis;
KW restenosis; autoimmune disorder; Alzheimer's disease;
KW peripheral neuropathy; trauma; spinal cord injury; allergy;
KW hematopoietic disorder; skeletal disorder; neurological disorder;
KW arthritic disorder; asthma; immunodeficiency disease; AIDS;
KW transplant rejection; ss.
OS Homo sapiens.
PN WO9911293-A1.
PD 11-MAR-1999.
PF 03-SEP-1998; U18360.
PR 12-SEP-1997; US-058974.
PR 05-SEP-1997; US-057626.
PR 05-SEP-1997; US-057663.
PR 05-SEP-1997; US-057669.
PR 12-SEP-1997; US-058666.
PR 12-SEP-1997; US-058667.
PR 12-SEP-1997; US-058973.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Brewer LA, Ebner R, Lafleur DW, Moore PA, Olsen HS,
PI Rosen GA, Ruben SM, Shi Y;
DR WPI; 99-204988/17.
DR P-PSDB; Y12941.
PT New isolated human genes and the secreted polypeptides they encode
PT - useful for diagnosis and treatment of e.g. neurological disorders,
PT tumours, immune disorders, inflammation or haematological disorders
PS Claim 1; Page 169; 215pp; English.
CC X51701-55 encode human secreted proteins. The polynucleotides and
CC their corresponding secreted polypeptides are useful for preventing,
CC treating or ameliorating medical conditions, e.g. by protein or gene
CC therapy. Pathological conditions can also be diagnosed by determining
CC the amount of the new polypeptides in a sample or by determining the
CC presence of mutations in the new polynucleotides. Specific uses are
CC described for each polynucleotide, based on which tissues they are
CC most highly expressed in, and include developing products for the
CC diagnosis or treatment of cancer, immune disorders, infection,
CC inflammatory disorders, skin disorders, tumours, atherosclerosis,
CC restenosis, autoimmune disorders, Alzheimer's disease, peripheral
CC neuropathies, trauma, spinal cord injuries, allergy, hematopoietic
CC disorders, skeletal disorders, neurological disorders, arthritic
CC disorders, asthma, immunodeficiency diseases, AIDS and transplant
CC rejection. The polypeptides are also useful for identifying their
CC binding partners.
SQ Sequence 921 BP; 250 A; 218 C; 174 G; 279 T;

Query Match 3.3%; Score 35.8; DB 1; Length 921;
Best Local Similarity 69.0%; Pred. No. 0.98;
Matches 49; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

```
Qy    1030 gcctggctttcatagtgaggctcagtgtcgaattaaagaggtaaagcaactattaaaaaa 1089
           ||||||  |   | || ||| ||| || || ||  ||| |   |||  ||  |  ||||||
Db     850 GCCTGGGTGATAAAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 909

Qy    1090 aaaaaaaaaaa 1100
           ||||||||||||
Db     910 AAAAAAAAAAA 920
```

RESULT 15
X37460
ID X37460 standard; cDNA; 1089 BP.
AC X37460;
DT 06-JUL-1999 (first entry)
DE Human secreted protein cDNA fragment containing gene 10.
KW Human; secreted protein; treatment; prevention; protein therapy; AIDS;
KW gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
KW developmental abnormality; fetal deficiency; blood disorder; leukemia;
KW immune system disease; autoimmune disease; hepatic disease; lymphoma;
KW renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;
KW cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;
KW pulmonary disorder; transplant rejection; osteoclast; osteoporosis;
KW arthritis; malignancy; digestive; endocrine; infection; ss.
OS Homo sapiens.
PN WO9918208-A1.
PD 15-APR-1999.
PF 01-OCT-1998; U20775.
PR 02-OCT-1997; US-060884.
PR 02-OCT-1997; US-060833.
PR 02-OCT-1997; US-060836.
PR 02-OCT-1997; US-060837.
PR 02-OCT-1997; US-060838.
PR 02-OCT-1997; US-060839.
PR 02-OCT-1997; US-060843.
PR 02-OCT-1997; US-060862.
PR 02-OCT-1997; US-060866.
PR 02-OCT-1997; US-060874.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Carter KC, Duan DR, Endress GA, Feng P, Ferrie AM,

PI Florence KA, Greene JM, Janat F, Lafleur DW, Ni J,
PI Rosen CA, Ruben SM, Shi Y, Young P, Yu G;
DR WPI; 99-264022/22.
DR P-PSDB; Y07861.
PT New isolated human genes and the secreted polypeptides they encode
PS Claim 1a; Page 232-233; 368pp; English.
CC This invention describes novel isolated human genes and the secreted
CC proteins they encode. The products of the invention are useful for
CC preventing, treating or ameliorating medical conditions, e.g. by protein
CC or gene therapy. Also pathological conditions can be diagnosed by
CC determining the amount of the new polypeptides in a sample or by
CC determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the 101 polynucleotides, based on
CC which tissues they are most highly expressed in, and include developing
CC products for the diagnosis or treatment of cancer, tumours,
CC neurodegenerative disorders, developmental abnormalities and fetal
CC deficiencies, blood disorders, leukemias, diseases of the immune system,
CC autoimmune diseases, hepatic and renal disease, lymphomas, inflammation,
CC allergies, Alzheimer's and cognitive disorders, schizophrenia, prostate
CC disease, skeletal or cardiac muscle disorders, pulmonary disorders,
CC transplant rejection, disorders involving osteoclasts such as
CC osteoporosis, arthritis or malignancies, digestive/endocrine disorders,
CC infections and AIDS. The human secreted proteins of the invention are
CC represented in Y07852-Y07993 and the encoding nucleic acids are
CC represented in X37451-X37552.
SQ Sequence 1089 BP; 367 A; 202 C; 240 G; 279 T;

Query Match 3.2%; Score 35.4; DB 1; Length 1089;
Best Local Similarity 69.6%; Pred. No. 1.4;
Matches 48; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

```
Qy   1030 gcctggctttcatagtgaggctcagtgtcgaattaaagaggtaaagcaactattaaaaaa 1089
          |||||| |  ||  ||||| ||| || || ||  ||| |   |||  ||  |  ||||||
Db   1015 GCCTGGGTGACAGGGTGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1074

Qy   1090 aaaaaaaaa 1098
          |||||||||
Db   1075 AAAAAAAAA 1083
```

Search completed: May 15, 2000, 01:45:56
Job time: 9820 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: May 14, 2000, 22:48:01 ; Search time 1646.42 Seconds
(without alignments)
-649.938 Million cell updates/sec

Title: US-09-223-796-1
Perfect score: 1100
Sequence: 1 gcacgagccacagccagcta.....attaaaaaaaaaaaaaaaa 1100

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vi:*
17: em_fun:*
18: em_hum1:*
19: em_hum2:*
20: em_in:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pl:*
27: em_ro:*
28: em_sts:*
29: em_sy:*
30: em_un:*
31: em_vi:*
32: gb_htg1:*
33: gb_htg2:*
34: gb_in1:*
35: gb_in2:*
36: em_ba1:*
37: em_ba2:*
38: em_hum3:*
39: em_hum4:*
40: gb_pr4:*
41: gb_htg3:*
42: gb_htg4:*
43: gb_htg5:*
44: gb_htg6:*
45: gb_htg7:*
46: em_htg1:*
47: em_htg2:*
48: em_htg3:*
49: em_hum5:*
50: gb_pl3:*
51: gb_pr5:*
52: gb_htg8:*
53: gb_htg9:*
54: gb_htg10:*
55: gb_htg11:*
56: gb_htg12:*
57: gb_htg13:*
58: gb_htg14:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	471.2	42.8	108369	40	AF124523	AF124523 Homo sapi
c	2	471.2	42.8	174923	52	AF216665	AF216665 Homo sapi
	3	468	42.5	69528	40	AF146367	AF146367 Homo sapi
	4	466.4	42.4	128009	40	AF118808	AF118808 Homo sapi
c	5	418.2	38.0	133200	43	AC011951	AC011951 Homo sapi
	6	412.6	37.5	1588	40	AF113540	AF113540 Homo sapi
	7	51.6	4.7	78543	44	AC011335	AC011335 Homo sapi
	8	50.2	4.6	78651	12	AC007636	AC007636 Mus muscu
	9	50.2	4.6	78783	12	AC007306	AC007306 Mus muscu
	10	48.6	4.4	138608	56	AC022781	AC022781 Mus muscu
	11	48.4	4.4	167859	55	AC016365	AC016365 Mus muscu
	12	47.6	4.3	230475	42	AC015891	AC015891 Mus muscu
c	13	46.5	4.2	238554	55	AC020965	AC020965 Mus muscu
	14	45.8	4.2	963	12	AF213391	AF213391 Mus muscu
c	15	45.6	4.1	7218	5	I66494	I66494 Sequence 14
	16	45.6	4.1	247428	52	AC021632	AC021632 Mus muscu
c	17	45	4.1	139097	33	AC004591	AC004591 Mus muscu
	18	45	4.1	171952	12	AC002324	AC002324 Mus muscu
	19	44.8	4.1	82098	40	AC006252	AC006252 Homo sapi
c	20	44.8	4.1	207032	45	AC012224	AC012224 Homo sapi
	21	44.6	4.1	172463	41	AC009948	AC009948 Homo sapi
	22	44.6	4.1	182180	44	AC011404	AC011404 Homo sapi
	23	44.6	4.1	191221	54	AC021629	AC021629 Mus muscu
c	24	44.6	4.1	233994	52	AC021627	AC021627 Mus muscu
	25	44.4	4.0	934	12	MMU17267	U17267 Mus musculu
c	26	44.4	4.0	230475	42	AC015891	AC015891 Mus muscu
	27	44.2	4.0	124479	55	AC020966	AC020966 Mus muscu
c	28	44.2	4.0	178964	43	AC012297	AC012297 Mus muscu
c	29	44.2	4.0	287292	55	AC020971	AC020971 Mus muscu
	30	44	4.0	13390	12	AF030884	AF030884 Mus muscu
c	31	44	4.0	141662	55	AC006447	AC006447 Mus muscu
	32	44	4.0	143586	55	AC006404	AC006404 Mus muscu
c	33	44	4.0	162038	53	AC019082	AC019082 Homo sapi
c	34	43.8	4.0	67473	55	AC023197	AC023197 Mus muscu
	35	43.8	4.0	173740	55	AC020964	AC020964 Mus muscu
	36	43.8	4.0	251364	12	MMAE000664	AE000664 Mus muscu
c	37	43.6	4.0	5006	12	MMU24171	U24171 Mus musculu
	38	43.6	4.0	238554	55	AC020965	AC020965 Mus muscu
	39	43.6	4.0	270294	55	AC020969	AC020969 Mus muscu
	40	43.4	3.9	267	12	DNY16208	Y16208 Dryomys nit
	41	43.4	3.9	189750	46	AC009860	Ac009860 Homo sapi
	42	43.4	3.9	206383	12	AC008160	AC008160 Mus muscu
c	43	43.2	3.9	269797	55	AC020968	AC020968 Mus muscu
	44	43	3.9	37349	12	MUSXRCC1G	L34078 Mouse XRCC1
	45	42.8	3.9	230383	55	AC021063	AC021063 Mus muscu

ALIGNMENTS

```
RESULT   1
AF124523/c
LOCUS       AF124523   108369 bp    DNA             PRI       24-AUG-1999
DEFINITION  Homo sapiens chromosome 8 clone Kox18-21 containing zinc-finger
            protein 7 (ZFP7) gene, complete sequence.
ACCESSION   AF124523
VERSION     AF124523.1  GI:4325309
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 108369)
  AUTHORS   Schudy,A., Rump,A., Drescher,B., Koczan,D., Thiesen,H.-J.,
            Merck-Rousseau,M.F. and Rosenthal,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-1999) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES             Location/Qualifiers
     source          1. .108369
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /clone="Kox18-21"
     repeat_region   1. .245
                     /rpt_family="AluY"
                     /evidence=not_experimental
     repeat_region   246. .584
                     /rpt_family="L1MA4"
                     /evidence=not_experimental
     repeat_region   585. .863
                     /rpt_family="AluJo"
                     /evidence=not_experimental
     repeat_region   864. .1068
                     /rpt_family="L1MA4"
                     /evidence=not_experimental
     repeat_region   1069. .1353
                     /rpt_family="AluSx"
                     /evidence=not_experimental
     repeat_region   1387. .1734
                     /rpt_family="L1MA4"
                     /evidence=not_experimental
     repeat_region   1735. .2042
                     /rpt_family="AluY"
                     /evidence=not_experimental
     repeat_region   2043. .2172
                     /rpt_family="L1MA4"
                     /evidence=not_experimental
     repeat_region   2173. .2482
                     /rpt_family="AluSx"
                     /evidence=not_experimental
     repeat_region   2483. .2901
                     /rpt_family="L1MA4"
                     /evidence=not_experimental
     exon            2835. .2912
                     /note="GRAIL, score = 45.000%, comment = marginal"
                     /evidence=not_experimental
     exon            2848. .2883
                     /note="Xpound exon prediction, score = 77% (0%)"
                     /evidence=not_experimental
     repeat_region   2902. .3202
                     /rpt_family="AluSp"
                     /evidence=not_experimental
     repeat_region   3203. .3325
                     /rpt_family="L1MA4"
                     /evidence=not_experimental
     repeat_region   3326. .3625
                     /rpt_family="AluSx"
                     /evidence=not_experimental
     repeat_region   3626. .3968
                     /rpt_family="L1MA4"
                     /evidence=not_experimental
     repeat_region   4104. .4431
                     /rpt_family="L2"
                     /evidence=not_experimental
     exon            4123. .4369
                     /note="GRAIL, score = 63.000%, comment = good"
                     /evidence=not_experimental
     repeat_region   4450. .4608
                     /rpt_family="AluSg/x"
                     /evidence=not_experimental
     repeat_region   4623. .4796
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   4898. .4997
                     /rpt_family="AluSg"
                     /evidence=not_experimental
     repeat_region   5000. .5060
                     /rpt_family="L1MA4A"
                     /evidence=not_experimental
     repeat_region   5101. .5402
                     /rpt_family="AluY"
                     /evidence=not_experimental
     repeat_region   complement(5578. .5595)
                     /rpt_family="MER33"
                     /evidence=not_experimental
     repeat_region   5596. .5867
                     /rpt_family="AluSx"
                     /evidence=not_experimental
     repeat_region   5902. .5912
                     /rpt_family="AluSx"
                     /evidence=not_experimental
     repeat_region   complement(5913. .6139)
                     /rpt_family="MER33"
                     /evidence=not_experimental
     repeat_region   complement(8764. .9075)
                     /rpt_family="AluSx"
                     /evidence=not_experimental
     repeat_region   10935. .11017
                     /rpt_family="MER74B"
                     /evidence=not_experimental
     repeat_region   complement(11018. .11306)
                     /rpt_family="AluJb"
                     /evidence=not_experimental
     repeat_region   11307. .11724
                     /rpt_family="MER74B"
                     /evidence=not_experimental
     repeat_region   11725. .12006
                     /rpt_family="AluSq"
                     /evidence=not_experimental
     repeat_region   12049. .12100
                     /note="homology = 100.00%, score = 26, counts = 2"
                     /rpt_type=tandem
                     /rpt_unit=actcaccttgagatgtgggtgtctgc
                     /evidence=not_experimental
     exon            complement(12081. .12230)
                     /note="GRAIL, score = 56.000%, comment = good"
                     /evidence=not_experimental
     repeat_region   complement(12255. .12748)
                     /rpt_family="LTR42"
                     /evidence=not_experimental
     exon            14255. .14377
                     /note="GenScan, score = 10.27%, comment = Initial_exon 123
                     bp frame: 1 phase: 0"
                     /evidence=not_experimental
     exon            14386. .14580
                     /gene="zinc-finger protein 7 (ZFP7)"
                     /number=1
     mRNA            join(14386. .14580,15811. .15858,16246. .16372,23346. .23462,
                     27285. .29150)
                     /gene="zinc-finger protein 7 (ZFP7)"
                     /note="mRNA acc.no. M29580"
                     /product="ZFP7"
     gene            14386. .29150
                     /gene="zinc-finger protein 7 (ZFP7)"
```

```
     repeat_region   15176. .15595
                     /rpt_family="L2"
                     /evidence=not_experimental
     exon            15811. .15858
                     /gene="zinc-finger protein 7 (ZFP7)"
                     /number=2
     CDS             join(15856. .15858,16246. .16372,23346. .23462,27285. .29098)
                     /gene="zinc-finger protein 7 (ZFP7)"
                     /note="mRNA acc.no. M29580"
                     /codon_start=1
                     /product="zinc-finger protein 7"
                     /protein_id="AAD17326.1"
                     /db_xref="GI:4325310"
                     /translation="MEVVTFGDVAVHFSREEWQCLDPGQRALYREVMLENHSSVAGLA
                     GFLVFKPELISRLEQGEEPWVLDLQGAEGTEAPRTSKTDSTIRTENEQACEDMDILKS
                     ESYGTVVRISPQDFPQNPGFGDVSDSEVWLDSHLGSPGLKVTGFTFQNNCLNEETVVP
                     KTFTKDAPQGCKELGSSGLDCQPLESQGESAEGMSQRCEECGKGIRATSDIALHWEIN
                     TQKISRCQECQKKLSDCLQGKHTNNCHGEKPYECAECGKVFRLCSQLNQHQRIHTGEK
                     PFKCTECGKAFRLSSKLIQHQRIHTGEKPYRCEECGKAFGQSSSLIHHQRIHTGERPY
                     GCRECGKAFSQQSQLVRHQRTHTGERPYPCKECGKAFSQSSTLAQHQRMHTGEKAQIL
                     KASDSPSLVAHQRIHAVEKPFKCDECGKAFRWISRLSQHQLIHTGEKPYKCNKCTKAF
                     GCSSRLIRHQRTHTGEKPFKCDECGKGFVQGSHLIQHQRIHTGEKPYVCNDCGKAFSQ
                     SSSLIYHQRIHKGEKPYECLQCGKAFSMSTQLTIHQRVHTGERPYKCNECGKAFSQNS
                     TLFQHQIIHAGVKPYECSECGKAFSRSSYLIEHQRIHTRAQWFYEYGNALEGSTFVSR
                     KKVNTIKKLHQCEDCEKIFRWRSHLIIHQRIHTGEKPYKCNDCGKAFNRSSRLTQHQK
                     IHMG"
     exon            16246. .16372
                     /gene="zinc-finger protein 7 (ZFP7)"
                     /number=3
     repeat_region   complement(17102. .17300)
                     /rpt_family="L1ME"
                     /evidence=not_experimental
     repeat_region   complement(17497. .17665)
                     /rpt_family="AluSg/x"
                     /evidence=not_experimental
     repeat_region   18270. .18401
                     /rpt_family="AluYb8"
                     /evidence=not_experimental
     repeat_region   18404. .18692
                     /rpt_family="AluSx"
                     /evidence=not_experimental
     repeat_region   18871. .19172
                     /rpt_family="AluSx"
                     /evidence=not_experimental
     repeat_region   complement(19300. .19632)
                     /rpt_family="AluSx"
                     /evidence=not_experimental
     repeat_region   19635. .19688
                     /note="homology = 74.10%, score = 20, counts = 9"
                     /rpt_type=tandem
                     /rpt_unit=tttgtt
                     /evidence=not_experimental
     repeat_region   complement(19740. .20031)
                     /rpt_family="AluSp"
                     /evidence=not_experimental
     repeat_region   complement(20033. .20103)
                     /rpt_family="L1M4"

  Query Match             42.8%;  Score 471.2;  DB 40;  Length 108369;
  Best Local Similarity   77.3%;  Pred. No. 7.6e-126;
  Matches  572;  Conservative    0;  Mismatches  168;  Indels    0;  Gaps    0;

Qy      84 tgtggttaggaccttctcttcggactggtcaagaaacgggaagaaaggatgtctgctttg 143
           | |||    ||||| ||   | | ||| ||||  |    | || |  |||||||||| ||
Db   37314 TCTGGGACCGACCTCCTGGGCTGGCTGATCAAAGAGGAAGCAGCAGCAATGTCTGCTGTG 37255

Qy     144 ggggctgcagctccatacttgcaccatcccgctgacagtcacagtggccgggtcagtttc 203
           |||||||||| |||||||| |||| ||||| | ||| ||||||||||||||| || ||||||
Db   37254 GGGGCTGCAACTCCATACCTGCATCATCCTGGTGATAGTCACAGTGGCCGAGTGAGTTTC 37195

Qy     204 ctgggttcccagccctctccagaagtgacggccgtggctcagctcttgaaggacttagac 263
           ||||  ||||||   ||||||| ||| | ||  |||| | |||  |   |||| |||||
Db   37194 TTGGGGGCCCAGCTTCCTCCAGAGGTGGCAGCAATGGCCCGGCTACTAGGGGACCTAGAC 37135

Qy     264 aggagcaccttcagaaagttgttgaaacttgtagtcggggccctgcatgggaaagactgc 323
           |||||||| |||||||||||| |||| |||| ||| |   |||||| ||| | ||||||
Db   37134 AGGAGCACGTTCAGAAAGTTGCTGAAGTTTGTGGTCAGCAGCCTGCAGGGGGAGGACTGC 37075

Qy     324 agagaagctgtggagcaacttggtgccagcgccaacctgtcagaagagcgtctggccgtc 383
           |||| |||||| ||| |||| | |||||||||||||| | || |||| |||| | |
Db   37074 CGAGAGGCTGTGCAGCGTCTTGGGGTCAGCGCCAACCTGCCGGAGGAGCAGCTGGGTGCC 37015

Qy     384 ctgctggcgggcacacacaccctgctccagcaggctctccggctgccccctgctagtcta 443
           |||||||| |||| ||||| |||||||||||||| ||||| |||||||| | || ||
Db   37014 CTGCTGGCAGGCATGCACACACTGCTCCAGCAGGCCCTCCGTCTGCCCCCCACCAGCCTG 36955

Qy     444 aagccagatgccttccaggaagagctccaggaacttggcattcctcaggatctaattgga 503
           ||||| ||  ||||| ||| |||||||||| ||  |||| || || || ||  | ||
Db   36954 AAGCCTGACACCTTCAGGGACCAGCTCCAGGAGCTCTGCATCCCCCAAGACCTGGTCGGG 36895

Qy     504 gatttggccagtttggcatttgggagtcaacgccctcttctcgactctgtagcccaacag 563
           || |||||||| ||| |||||||||| || || || || || || || ||||| ||||| |||
Db   36894 GACTTGGCCAGCGTGGTATTTGGGAGCCAGCGGCCCCTCCTTGATTCTGTGGCCCAGCAG 36835

Qy     564 cagggatcctcgctgcctcacgtgtcttacttccggtggcgggtggacgtggccatctca 623
           |||||  ||| |||||| || ||  || |||| |||||||||||||| || || |||||
Db   36834 CAGGGGGCCTGGCTGCCGCATGTTGCTGACTTTCGGTGGCGGGTGGATGTAGCAATCTCC 36775

Qy     624 accagcgctcagtcccgctccctgcaaccgagtgttctcatgcagctgaagctcacagat 683
           ||||| || | | | |||||||||||| ||||| || || ||||||||||||||  |||||
Db   36774 ACCAGTGCCCTGGCTCGCTCCCTGCAGCCGAGCGTCCTGATGCAGCTGAAGCTTTCAGAT 36715

Qy     684 ggatctgcacaccgcttcgaggtgcccatagccaaatttcaggagctgcggtacagtgta 743
           || || ||| ||||||| ||||| |||| |||||| || ||||||||||||||||| ||
Db   36714 GGGTCAGCATACCGCTTTGAGGTCCCCACAGCCAAGTTCCAGGAGCTGCGGTACAGCGTG 36655

Qy     744 gccttggtccttaaggagatggcagaactggagaagaagtgtgagcgcaaactgcaggac 803
           ||| ||||||| |||||||||||||| |||||||||| |||||||||| |||||||||
Db   36654 GCCCTGGTCCTAAAGGAGATGGCAGATCTGGAGAAGAGGTGTGAGCGCAGACTGCAGGAC 36595

Qy     804 tgactgaaccctggtactgt 823
           ||||    | || |  | ||
Db   36594 TGACCCCTCACTTGACCAGT 36575


RESULT   2
AF216665/c
LOCUS       AF216665   174923 bp    DNA             HTG       30-JAN-2000
DEFINITION  Homo sapiens chromosome 8 clone BAC 704B241056 map 8q24, ***
            SEQUENCING IN PROGRESS ***, in unordered pieces.
ACCESSION   AF216665
VERSION     AF216665.1  GI:6815808
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 174923)
  AUTHORS   Schudy,A., Schilhabel,M., Baumgart,C., Menzel,U., Weber,J.,
            Schattevoy,R. and Rosenthal,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-DEC-1999) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
COMMENT     1-75864: contig of 75864 bp;  75864-75865: gap of unknown size;
            75865-109038: contig of 33174 bp;  109038-109039: gap of unknown
            size;  109039-117368: contig of 8330 bp;  117368-117369: gap of
            unknown size;  117369-128688: contig of 11320 bp;  128688-128689:
            gap of unknown size;  128689-134145: contig of 5457 bp;
            134145-134146: gap of unknown size;  134146-147620: contig of 13475
            bp;  147620-147621: gap of unknown size;  147621-149478: contig of
            1858 bp;  149478-149479: gap of unknown size;  149479-151777:
            contig of 2299 bp;  151777-151778: gap of unknown size;
            151778-155130: contig of 3353 bp;  155130-155131: gap of unknown
            size;  155131-157254: contig of 2124 bp;  157254-157255: gap of
            unknown size;  157255-158492: contig of 1238 bp;  158492-158493:
            gap of unknown size;  158493-174923: contig of 16431 bp;.
```

```
          * NOTE: This is a 'working draft' sequence.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
FEATURES             Location/Qualifiers
     source          1. .174923
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8q24"
                     /clone="BAC 704B241056"
     repeat_region   complement(36. .349)
                     /rpt_family="AluSx"
     repeat_region   352. .405
                     /rpt_type=tandem
     exon            complement(449. .582)
                     /note="FEXHB"
                     /evidence=not_experimental
     repeat_region   complement(457. .748)
                     /rpt_family="AluSq"
     repeat_region   complement(750. .820)
                     /rpt_family="L1M4_orf2"
     repeat_region   complement(1167. .1754)
                     /rpt_family="L1M4_orf2"
     repeat_region   1387. .1410
                     /rpt_type=tandem
     repeat_region   complement(1410. .1713)
                     /rpt_family="AluY"
     exon            1434. .1555
                     /note="FEXHB"
                     /evidence=not_experimental
     repeat_region   1517. .1702
                     /rpt_type=inverted
     repeat_region   complement(1865. .2170)
                     /rpt_family="AluJo"
     repeat_region   2386. .2461
                     /rpt_type=tandem
     repeat_region   complement(2435. .2544)
                     /rpt_family="AluJb"
     repeat_region   complement(2545. .2649)
                     /rpt_family="AluSx"
     repeat_region   2738. .3017
                     /rpt_family="AluJo"
     repeat_region   2747. .2931
                     /rpt_type=inverted
     exon            3369. .3458
                     /note="GENSCAN"
     exon            4066. .4182
                     /note="GENSCAN"
     exon            4066. .4182
                     /note="MZEF"
                     /evidence=not_experimental
     repeat_region   complement(4603. .5038)
                     /rpt_family="MLT1C"
     exon            complement(4836. .4889)
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            complement(5199. .5235)
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            complement(5199. .5235)
                     /note="MZEF"
                     /evidence=not_experimental
     repeat_region   5336. .5612
                     /rpt_family="AluY"
     repeat_region   5472. .5530
                     /rpt_type=inverted
     repeat_region   5627. .5906
                     /rpt_type=tandem
     repeat_region   complement(6589. .6751)
                     /rpt_family="AluSg"
     repeat_region   6689. .6747
                     /rpt_type=inverted
     exon            8005. .8268
                     /note="MZEF"
                     /evidence=not_experimental
     exon            8005. .9818
                     /note="GENSCAN"
     exon            8046. .9818
                     /note="FEXHB"
                     /evidence=not_experimental
     exon            8046. .9818
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            complement(9207. .9412)
                     /note="GRAIL"
                     /evidence=not_experimental
     repeat_region   9299. .9578
                     /rpt_type=tandem
     repeat_region   9958. .10259
                     /rpt_family="AluY"
     exon            complement(10113. .10232)
                     /note="FEXHB"
                     /evidence=not_experimental
     repeat_region   10238. .10259
                     /rpt_type=tandem
     repeat_region   11078. .11384
                     /rpt_family="AluY"
     repeat_region   11359. .11384
                     /rpt_type=tandem
     repeat_region   11519. .11638
                     /rpt_type=tandem
     repeat_region   complement(11792. .11848)
                     /rpt_family="LTR40b"
     exon            complement(11834. .11860)
                     /note="GENSCAN"
     repeat_region   complement(12208. .12332)
                     /rpt_family="HERVL40"
     exon            complement(12368. .12424)
                     /note="GENSCAN"
     exon            complement(12368. .12424)
                     /note="XPOUND"
                     /evidence=not_experimental
     exon            complement(12368. .12468)
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            complement(12673. .12773)
                     /note="GRAIL"
                     /evidence=not_experimental
     repeat_region   12872. .12961
                     /rpt_family="HERVL40"
     repeat_region   12985. .13069
                     /rpt_family="MER58B"
     repeat_region   13354. .13491
                     /rpt_type=inverted
     repeat_region   complement(13354. .13641)
                     /rpt_family="AluY"
     repeat_region   13949. .14111
                     /rpt_family="FRAM"
     repeat_region   14263. .14357
                     /rpt_family="L2b"
     exon            complement(14534. .14650)
                     /note="GRAIL"
                     /evidence=not_experimental
     repeat_region   14715. .14969
                     /rpt_family="AluJo"
     repeat_region   15002. .15162
                     /rpt_family="AluSq"
     repeat_region   15022. .15161
                     /rpt_type=inverted
     repeat_region   15168. .15812
                     /rpt_family="MER58A"
     repeat_region   16383. .16634
                     /rpt_family="L2"
     exon            16705. .16831
                     /note="GRAIL"
```

```
                  /evidence=not_experimental
     repeat_region   16714. .16761
                  /rpt_family="L2a"
     repeat_region   16722. .16761
                  /rpt_family="L2b"
     repeat_region   16782. .16912
                  /rpt_family="L2a"
     exon            complement(17315. .17989)
                  /note="GRAIL"
                  /evidence=not_experimental
     exon            complement(17315. .18046)
                  /note="GENSCAN"
     exon            complement(17321. .17848)
                  /note="XPOUND"
                  /evidence=not_experimental
     exon            complement(17338. .17681)
                  /note="M2EF"
                  /evidence=not_experimental
     exon            17351. .17527
                  /note="GRAIL"
                  /evidence=not_experimental
     exon            17537. .17863
                  /note="GRAIL"
                  /evidence=not_experimental
     exon            complement(18086. .18181)

  Query Match             42.8%;  Score 471.2;  DB 52;  Length 174923;
  Best Local Similarity   77.3%;  Pred. No. 7.9e-126;
  Matches  572;  Conservative    0;  Mismatches  168;  Indels    0;  Gaps    0;

Qy      84 tgtggttaggaccttctcttcggactggtcaagaaacgggaagaaaggatgtctgctttg 143
           | |||    ||||| ||   | | ||| |||| |    | || |   |||||||||| ||
Db   18037 TCTGGGACCGACCTCCTGGGCTGGCTGATCAAAGAGGAAGCAGCAGCAATGTCTGCTGTG 17978

Qy     144 ggggctgcagctccatacttgcaccatcccgctgacagtcacagtggccgggtcagtttc 203
           |||||||||| |||||||| |||| ||||| | ||| |||||||||||||| || ||||||
Db   17977 GGGGCTGCAACTCCATACCTGCATCATCCTGGTGATAGTCACAGTGGCCGAGTGAGTTTC 17918

Qy     204 ctgggttcccagccctctccagaagtgacggccgtggctcagctcttgaaggacttagac 263
            ||||  ||||||   ||||||| ||| | ||  |||| | |||  |   |||| |||||
Db   17917 TTGGGGGCCCAGCTTCCTCCAGAGGTGGCAGCAATGGCCCGGCTACTAGGGGACCTAGAC 17858

Qy     264 aggagcaccttcagaaagttgttgaaacttgtagtcggggccctgcatgggaaagactgc 323
           |||||||| |||||||||||| |||| |||| ||| | |||||| ||| | ||||||
Db   17857 AGGAGCACGTTCAGAAAGTTGCTGAAGTTTGTGGTCAGCAGCCTGCAGGGGGAGGACTGC 17798

Qy     324 agagaagctgtggagcaacttggtgccagcgccaacctgtcagaagagcgtctggccgtc 383
            |||| |||||| |||  ||||| | |||||||||||||| | || ||||  ||||  | |
Db   17797 CGAGAGGCTGTGCAGCGTCTTGGGGTCAGCGCCAACCTGCCGGAGGAGCAGCTGGGTGCC 17738

Qy     384 ctgctggcgggcacacacaccctgctccagcaggctctccggctgccccctgctagtcta 443
           |||||||| ||||  ||||| |||||||||||||||| ||||| ||||||||  | || ||
Db   17737 CTGCTGGCAGGCATGCACACACTGCTCCAGCAGGCCCTCCGTCTGCCCCCCACCAGCCTG 17678

Qy     444 aagccagatgccttccaggaagagctccaggaacttggcattcctcaggatctaattgga 503
           ||||| ||  |||||  |||  |||||||||| ||  |||| || || || ||  | ||
Db   17677 AAGCCTGACACCTTCAGGGACCAGCTCCAGGAGCTCTGCATCCCCCAAGACCTGGTCGGG 17618

Qy     504 gatttggccagtttggcatttgggagtcaacgccctcttctcgactctgtagcccaacag 563
           || ||||||||  ||| |||||||| || || || || || || ||||| ||||| |||
Db   17617 GACTTGGCCAGCGTGGTATTTGGGAGCCAGCGGCCCCTCCTTGATTCTGTGGCCCAGCAG 17558

Qy     564 cagggatcctcgctgcctcacgtgtcttacttccggtggcgggtggacgtggccatctca 623
           |||||  ||| |||||| || ||  || |||| ||||||||||||| || || |||||
Db   17557 CAGGGGGCCTGGCTGCCGCATGTTGCTGACTTTCGGTGGCGGGTGGATGTAGCAATCTCC 17498

Qy     624 accagcgctcagtcccgctccctgcaaccgagtgttctcatgcagctgaagctcacagat 683
           ||||| || | | | |||||||||||| ||||| || || ||||||||||||||  |||||
Db   17497 ACCAGTGCCCTGGCTCGCTCCCTGCAGCCGAGCGTCCTGATGCAGCTGAAGCTTTCAGAT 17438

Qy     684 ggatctgcacaccgcttcgaggtgcccatagccaaatttcaggagctgcggtacagtgta 743
           || || ||| ||||||| ||||| |||| |||||| || ||||||||||||||||  ||
Db   17437 GGGTCAGCATACCGCTTTGAGGTCCCCACAGCCAAGTTCCAGGAGCTGCGGTACAGCGTG 17378

Qy     744 gccttggtccttaaggagatggcagaactggagaagaagtgtgagcgcaaactgcaggac 803
           ||| |||||||| |||||||||||||| |||||||||| |||||||||| |||||||||
Db   17377 GCCCTGGTCCTAAAGGAGATGGCAGATCTGGAGAAGAGGTGTGAGCGCAGACTGCAGGAC 17318

Qy     804 tgactgaaccctggtactgt 823
           ||||    | || |  | ||
Db   17317 TGACCCCTCACTTGACCAGT 17298


RESULT   3
AF146367
LOCUS       AF146367    69528 bp    DNA             PRI       20-MAY-1999
DEFINITION  Homo sapiens chromosome 8 PAC P5A4.00 containing gene for zinc
            finger protein, complete sequence.
ACCESSION   AF146367
VERSION     AF146367.1  GI:4877991
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 69528)
  AUTHORS   Rump,A., Schudy,A., Drescher,B., Koczan,D., Thiessen,H.-J.,
            Merck-Rousseau,M.-F. and Rosenthal,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-APR-1999) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES             Location/Qualifiers
     source          1. .69528
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /clone="PAC P5A4.00"
     gene            5504. .6918
                     /gene="hypothetical"
                     /note="GenScan, score = 1186%, comment = P1_377_aa 1. .208
                     of owl|P15622|ZF64_HUMAN ZINC FINGER PROTEIN CLONE 647
                     (FRAGMENT). - HOMO SAPIENS P = 1.2e-120 S = 1186"
                     /evidence=not_experimental
     exon            5504. .6637
                     /gene="hypothetical"
                     /note="GenScan, score = 117.75%, comment =
                     Single-exon_gene 1134 bp frame: 1 phase: 0"
                     /evidence=not_experimental
     gene            7855. .44516
                     /gene="hypothetical"
                     /note="GenScan, score = 0%, comment = P2_471_aa GENSCAN
                     predicted peptide is unknown ->NN (DEF2) prediction for
                     sec struc composition: all beta sheet; overall class:
                     lectin"
                     /evidence=not_experimental
     exon            9896. .10003
                     /gene="hypothetical"
                     /note="GenScan, score = -0.43%, comment = Initial_exon 108
                     bp frame: 1 phase: 0"
                     /evidence=not_experimental
     exon            20029. .20093
                     /gene="hypothetical"
                     /note="GenScan, score = 0.51%, comment = Internal_exon 65
                     bp frame: 0 phase: 2"
                     /evidence=not_experimental
     exon            20121. .20324
                     /gene="hypothetical"
                     /note="GenScan, score = 0.84%, comment = Internal_exon 204
                     bp frame: 0 phase: 0"
                     /evidence=not_experimental
     exon            26797. .26852
                     /gene="hypothetical"
                     /note="GenScan, score = 0.81%, comment = Internal_exon 56
                     bp frame: 1 phase: 2"
                     /evidence=not_experimental
     exon            27802. .27935
```

```
                     /gene="hypothetical"
                     /note="GenScan, score = 4.67%, comment = Internal_exon 134
                     bp frame: 2 phase: 2"
                     /evidence=not_experimental
     exon            31554. .31670
                     /gene="hypothetical"
                     /note="GenScan, score = 0.77%, comment = Internal_exon 117
                     bp frame: 2 phase: 0"
                     /evidence=not_experimental
     exon            36757. .37488
                     /gene="hypothetical"
                     /note="GenScan, score = 71.21%, comment = Terminal_exon
                     732 bp frame: 0 phase: 0"
                     /evidence=not_experimental
     gene            complement(44951. .60537)
                     /gene="KOX4"
                     /note="GenScan, score = 3534%, comment = P3_742_aa 44. .686
                     of owl|P17097|ZN07_HUMAN ZINC FINGER PROTEIN 7 (ZINC
                     FINGER PROTEIN KOX4) (ZINC P = 0.0 S = 3534"
                     /evidence=not_experimental
     exon            complement(44982. .46795)
                     /gene="KOX4"
                     /note="GenScan, score = 113.55%, comment = Terminal_exon
                     1814 bp frame: 2 phase: 2"
                     /evidence=not_experimental
     exon            complement(50641. .50757)
                     /gene="KOX4"
                     /note="GenScan, score = 7.44%, comment = Internal_exon 117
                     bp frame: 1 phase: 0"
                     /evidence=not_experimental
     exon            complement(51365. .51454)
                     /gene="KOX4"
                     /note="GenScan, score = 1.36%, comment = Internal_exon 90
                     bp frame: 2 phase: 0"
                     /evidence=not_experimental
     exon            complement(56929. .57014)
                     /gene="KOX4"
                     /note="GenScan, score = 6.06%, comment = Internal_exon 86
                     bp frame: 1 phase: 2"
                     /evidence=not_experimental
     exon            complement(60537. .60658)
                     /note="GenScan, score = 10.62%, comment = Initial_exon 122
                     bp frame: 1 phase: 2"
                     /evidence=not_experimental
BASE COUNT    17500 a  18304 c  16802 g  16922 t
ORIGIN
```

Query Match 42.5%; Score 468; DB 40; Length 69528;
Best Local Similarity 77.0%; Pred. No. 6.2e-125;
Matches 570; Conservative 0; Mismatches 170; Indels 0; Gaps 0;

```
Qy      84 tgtggttaggaccttctcttcggactggtcaagaaacgggaagaaaggatgtctgctttg 143
           | |||    ||||| ||   | | ||| |||| |    | || |   |||||||||| ||
Db   36766 TCTGGGACCGACCTCCTGGGCTGGCTGATCAAAGAGGAAGCAGCAGCAATGTCTGCTGTG 36825

Qy     144 ggggctgcagctccatacttgcaccatcccgctgacagtcacagtggccgggtcagtttc 203
           ||| ||||| |||||||| |||| ||||| | ||| ||||||||||||||| || ||||||
Db   36826 GGGACTGCAACTCCATACCTGCATCATCCTGGTGATAGTCACAGTGGCCGAGTGAGTTTC 36885

Qy     204 ctgggttcccagccctctccagaagtgacggccgtggctcagctcttgaaggacttagac 263
            ||||  ||||||   ||||||| ||| | ||  |||| | |||  |   |||| |||||
Db   36886 TTGGGGGCCCAGCTTCCTCCAGAGGTGGCAGCAATGGCCCGGCTACTAGGGGACCTAGAC 36945

Qy     264 aggagcaccttcagaaagttgttgaaacttgtagtcggggccctgcatgggaaagactgc 323
           |||||||| |||||||||||| ||||  |||| ||| |    |||||| ||| | ||||||
Db   36946 AGGAGCACGTTCAGAAAGTTGCTGAAGTTTGTGGTCAGCAGCCTGCAGGGGGAGGACTGC 37005

Qy     324 agagaagctgtggagcaacttggtgccagcgccaacctgtcagaagagcgtctggccgtc 383
            |||| ||||||  |||  ||||| | |||||||||||||| | || ||||  ||||  | |
Db   37006 CGAGAGGCTGTGCAGCGTCTTGGGGTCAGCGCCAACCTGCCGGAGGAGCAGCTGGGTGCC 37065

Qy     384 ctgctggcgggcacacacaccctgctccagcaggctctccggctgccccctgctagtcta 443
           |||||||| ||||  ||||| |||||||||||||| ||||| ||||||||  | ||| ||
Db   37066 CTGCTGGCAGGCATGCACACACTGCTCCAGCAGGCCCTCCGTCTGCCCCCCACCAGCCTG 37125

Qy     444 aagccagatgccttccaggaagagctccaggaacttggcattcctcaggatctaattgga 503
           ||||| ||  |||||  |||  |||||||||| ||  |||| || || || || | ||
Db   37126 AAGCCTGACACCTTCAGGGACCAGCTCCAGGAGCTCTGCATCCCCCAAGACCTGGTCGGG 37185

Qy     504 gatttggccagtttggcatttgggagtcaacgccctcttctcgactctgtagcccaacag 563
           || ||||||||  ||| |||||||| || || || || || || ||||| ||||| |||
Db   37186 GACTTGGCCAGCGTGGTATTTGGGAGCCAGCGGCCCCTCCTTGATTCTGTGGCCCAGCAG 37245

Qy     564 cagggatcctcgctgcctcacgtgtcttacttccggtggcgggtggacgtggccatctca 623
           |||||  ||| |||||| || ||  || |||| |||||||||||||| || || || ||
Db   37246 CAGGGGGCCTGGCTGCCGCATGTTGCTGACTTTCGGTGGCGGGTGGATGTAGCAATTTCC 37305

Qy     624 accagcgctcagtcccgctccctgcaaccgagtgttctcatgcagctgaagctcacagat 683
           ||||| || | | | |||||||||| ||||| || || ||||||||||||||  |||||
Db   37306 ACCAGTGCCCTGGCTCGCTCCCTGCAGCCGAGCGTCCTGATGCAGCTGAAGCTTTCAGAT 37365

Qy     684 ggatctgcacaccgcttcgaggtgcccatagccaaatttcaggagctgcggtacagtgta 743
           || || ||| ||||||| ||||| |||| |||||| || ||||||||||||||||| ||
Db   37366 GGGTCAGCATACCGCTTTGAGGTCCCCACAGCCAAGTTCCAGGAGCTGCGGTACAGCGTG 37425

Qy     744 gccttggtccttaaggagatggcagaactggagaagaagtgtgagcgcaaactgcaggac 803
           ||| ||||||| |||||||||||||| |||||||||| |||||||||| ||||||||||
Db   37426 GCCCTGGTCCTAAAGGAGATGGCAGATCTGGAGAAGAGGTGTGAGCGCAGACTGCAGGAC 37485

Qy     804 tgactgaaccctggtactgt 823
           ||||    | || |  | ||
Db   37486 TGACCCCTCACTTGACCAGTT 37505
```

```
RESULT   4
AF118808
LOCUS       AF118808   128009 bp    DNA             PRI       30-JUN-1999
DEFINITION  Homo sapiens chromosome 8qter PAC clone P2G2, complete sequence.
ACCESSION   AF118808
VERSION     AF118808.1  GI:4337002
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 128009)
  AUTHORS   Rump,A., Drescher,B., Weber,J., Schattevoy,R.,
            Koczan,D., Thiesen,H.-J. and Merck-Rousseau,M.-F.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JAN-1999) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES             Location/Qualifiers
     source          1. .128009
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8qter"
                     /clone="PAC clone P2G2"
                     /note="mapped to 8qter by FISH; identified by filter
                     hybridization with zinc finger specific probe: clone
                     contains two sequencing gaps (60562-60589 and
                     68689-68687); orientation of the middle contig is not
                     known"
     exon            complement(2184. .2219)
                     /note="Exon predicted by several exon prediction programs
                     and by homology to GenBank entry HUMZIFIA_1: Homo sapiens
                     zinc finger protein mRNA, 3' end."
                     /evidence=not_experimental
     exon            2269. .2310
                     /note="Exon predicted by several exon prediction programs
                     and by homology to GenBank entry HUMZNFN_1: Homo sapiens
                     DNA-binding protein (ZNF) gene, partial cds."
                     /evidence=not_experimental
     exon            47097. .47263
                     /note="Exon predicted by several exon prediction programs
```

```
                     and by homology to GenBank entry HSKOX9_1: Human Kox9 mRNA
                     for zinc finger protein, partial."
                     /evidence=not_experimental
     exon            47430. .47631
                     /note="Exon predicted by several exon prediction programs
                     and by homology to GenBank entry AF024700_1: Homo sapiens
                     clone B4-2 zinc finger protein mRNA, partial cds."
                     /evidence=not_experimental
     exon            94802. .95455
                     /note="Exon predicted by several exon prediction programs
                     and by homology to GenBank entry HSZFP647_1: Human mRNA
                     for zinc finger protein (clone 647)."
                     /evidence=not_experimental
     exon            complement(95143. .95187)
                     /note="Exon predicted by several exon prediction programs
                     and by homology to GenBank entry HSZNFPT7_1: H.sapiens
                     mRNA for Zinc-finger protein (ZNFpT7)."
                     /evidence=not_experimental
BASE COUNT    35117 a  30514 c  28979 g  33352 t     47 others
ORIGIN


  Query Match           42.4%;  Score 466.4;  DB 40;  Length 128009;
  Best Local Similarity   76.9%;  Pred. No. 1.9e-124;
  Matches  569;  Conservative    0;  Mismatches  171;  Indels    0;  Gaps    0;

Qy      84 tgtggttaggaccttctcttcggactggtcaagaaacgggaagaaaggatgtctgctttg 143
           | |||    ||||| ||    | | ||| ||||  |    | || |   |||||||| ||
Db 125853 TCTGGGACCGACCTCCTGGGCTGGCTGATCAAAGAGGAAGCAGCAGCAATGTCTGCTGTG 125912

Qy     144 ggggctgcagctccatacttgcaccatcccgctgacagtcacagtggccgggtcagtttc 203
           |||||||||| |||||||| |||| ||||| | ||| |||||||||||||| || ||||||
Db 125913 GGGGCTGCAACTCCATACCTGCATCATCCTGGTGATAGTCACAGTGGCCGAGTGAGTTTC 125972

Qy     204 ctgggttcccagccctctccagaagtgacggccgtggctcagctcttgaaggacttagac 263
           ||||  ||||||   ||||| | ||| | ||  |||| | |||  |   |||| |||||
Db 125973 TTGGGGGCCCAGCTTCCTCCAAAGGTGGCAGCAATGGCCCGGCTACTAGGGGACCTAGAC 126032

Qy     264 aggagcaccttcagaaagttgttgaaacttgtagtcggggccctgcatgggaaagactgc 323
           |||||||| |||||||||||| |||| |||| ||| |   |||||| ||| | ||||||
Db 126033 AGGAGCACGTTCAGAAAGTTGCTGAAGTTTGTGGTCAGCAGCCTGCAGGGGGAGGACTGC 126092

Qy     324 agagaagctgtggagcaacttggtgccagcgccaacctgtcagaagagcgtctggccgtc 383
           || | |||||| | |  ||||| | |||||||||||||| | || ||||  ||||  | |
Db 126093 CGAAAGGCTGTGCACCGTCTTGGGGTCAGCGCCAACCTGCCGGAGGAGCAGCTGGGTGCC 126152

Qy     384 ctgctggcgggcacacacaccctgctccagcaggctctccggctgccccctgctagtcta 443
           |||||||| ||||  ||||| |||||||||||||||| ||||| ||||||||  | || ||
Db 126153 CTGCTGGCAGGCATGCACACACTGCTCCAGCAGGCCCTCCGTCTGCCCCCCACCAGCCTG 126212

Qy     444 aagccagatgccttccaggaagagctccaggaacttggcattcctcaggatctaattgga 503
           ||||| ||  |||||  |||  |||||||||| ||  |||| || || || ||  | ||
Db 126213 AAGCCTGACACCTTCAGGGACCAGCTCCAGGAGCTCTGCATCCCCCAAGACCTGGTCGGG 126272

Qy     504 gatttggccagtttggcatttgggagtcaacgccctcttctcgactctgtagcccaacag 563
           || |||||||||  ||| |||||||||| || || || || || || ||||| ||||| |||
Db 126273 GACTTGGCCAGCGTGGTATTTGGGAGCCAGCGGCCCCTCCTTGATTCTGTGGCCCAGCAG 126332

Qy     564 cagggatcctcgctgcctcacgtgtcttacttccggtggcgggtggacgtggccatctca 623
           |||||  ||| |||||| || ||  || |||| ||||||||||||||| || || |||||
Db 126333 CAGGGGGCCTGGCTGCCGCATGTTGCTGACTTTCGGTGGCGGGTGGATGTAGCAATCTCC 126392

Qy     624 accagcgctcagtcccgctccctgcaaccgagtgttctcatgcagctgaagctcacagat 683
           ||||| || | | |||||||||||| ||||| || || ||||||||||||||  |||||
Db 126393 ACCAGTGCCCTGGCTCGCTCCCTGCAGCCGAGCGTCCTGATGCAGCTGAAGCTTTCAGAT 126452

Qy     684 ggatctgcacaccgcttcgaggtgcccatagccaaatttcaggagctgcggtacagtgta 743
           || || ||| ||||||| ||||| |||| |||||| || ||||||||||||||||  ||
Db 126453 GGGTCAGCATACCGCTTTGAGGTCCCCACAGCCAAGTTCCAGGAGCTGCGGTACAGCGTG 126512

Qy     744 gccttggtccttaaggagatggcagaactggagaagaagtgtgagcgcaaactgcaggac 803
           ||| ||||||| |||||||||||||| |||||||||| |||||||||| ||||||||||
Db 126513 GCCCTGGTCCTAAAGGAGATGGCAGATCTGGAGAAGAGGTGTGAGCGCAGACTGCAGGAC 126572
```

```
Qy     804 tgactgaaccctggtactgt 823
           ||||    | || |  | ||
Db 126573 TGACCCCTCACTTGACCAGT 126592


RESULT   5
AC011951/c
LOCUS       AC011951   133200 bp    DNA             HTG       13-DEC-1999
DEFINITION  Homo sapiens clone RP11-18M17, *** SEQUENCING IN PROGRESS ***, 53
            unordered pieces.
ACCESSION   AC011951
VERSION     AC011951.2  GI:6554541
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 133200)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-18M17
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 133200)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Dec 10, 1999 this sequence version replaced gi:6056232.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L3701
                Center clone name: 18_M_17
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 53 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1041: contig of 1041 bp in length
            *                   gap of unknown length
            *     1042     1130: contig of 89 bp in length
            *                   gap of unknown length
            *     1131     2412: contig of 1282 bp in length
            *                   gap of unknown length
            *     2413     3457: contig of 1045 bp in length
            *                   gap of unknown length
            *     3458     4573: contig of 1116 bp in length
            *                   gap of unknown length
            *     4574     5792: contig of 1219 bp in length
            *                   gap of unknown length
```

```
*     5793     6853: contig of 1061 bp in length
*                  gap of unknown length
*     6854     8428: contig of 1575 bp in length
*                  gap of unknown length
*     8429     9591: contig of 1163 bp in length
*                  gap of unknown length
*     9592    10969: contig of 1378 bp in length
*                  gap of unknown length
*    10970    12475: contig of 1506 bp in length
*                  gap of unknown length
*    12476    14088: contig of 1613 bp in length
*                  gap of unknown length
*    14089    15243: contig of 1155 bp in length
*                  gap of unknown length
*    15244    16417: contig of 1174 bp in length
*                  gap of unknown length
*    16418    17553: contig of 1136 bp in length
*                  gap of unknown length
*    17554    18961: contig of 1408 bp in length
*                  gap of unknown length
*    18962    20529: contig of 1568 bp in length
*                  gap of unknown length
*    20530    23589: contig of 3060 bp in length
*                  gap of unknown length
*    23590    25249: contig of 1660 bp in length
*                  gap of unknown length
*    25250    26988: contig of 1739 bp in length
*                  gap of unknown length
*    26989    28426: contig of 1438 bp in length
*                  gap of unknown length
*    28427    29981: contig of 1555 bp in length
*                  gap of unknown length
*    29982    31552: contig of 1571 bp in length
*                  gap of unknown length
*    31553    33219: contig of 1667 bp in length
*                  gap of unknown length
*    33220    35694: contig of 2475 bp in length
*                  gap of unknown length
*    35695    37968: contig of 2274 bp in length
*                  gap of unknown length
*    37969    39872: contig of 1904 bp in length
*                  gap of unknown length
*    39873    42198: contig of 2326 bp in length
*                  gap of unknown length
*    42199    43732: contig of 1534 bp in length
*                  gap of unknown length
*    43733    45927: contig of 2195 bp in length
*                  gap of unknown length
*    45928    47914: contig of 1987 bp in length
*                  gap of unknown length
*    47915    50000: contig of 2086 bp in length
*                  gap of unknown length
*    50001    51834: contig of 1834 bp in length
*                  gap of unknown length
*    51835    54069: contig of 2235 bp in length
*                  gap of unknown length
*    54070    56363: contig of 2294 bp in length
*                  gap of unknown length
*    56364    59354: contig of 2991 bp in length
*                  gap of unknown length
*    59355    63071: contig of 3717 bp in length
*                  gap of unknown length
*    63072    65969: contig of 2898 bp in length
*                  gap of unknown length
*    65970    69646: contig of 3677 bp in length
*                  gap of unknown length
*    69647    73243: contig of 3597 bp in length
*                  gap of unknown length
*    73244    76844: contig of 3601 bp in length
*                  gap of unknown length
*    76845    79797: contig of 2953 bp in length
*                  gap of unknown length
*    79798    83127: contig of 3330 bp in length
*                  gap of unknown length
*    83128    85980: contig of 2855 bp in length
*                  gap of unknown length
*    85983    89291: contig of 3309 bp in length
*                  gap of unknown length
*    89292    92357: contig of 3066 bp in length
*                  gap of unknown length
*    92358    95772: contig of 3415 bp in length
*                  gap of unknown length
*    95773    99411: contig of 3639 bp in length
*                  gap of unknown length
*    99412   103446: contig of 4035 bp in length
*                  gap of unknown length
*   103447   108787: contig of 5341 bp in length
*                  gap of unknown length
*   108788   116339: contig of 7552 bp in length
*                  gap of unknown length
*   116340   124476: contig of 8137 bp in length
*                  gap of unknown length
*   124477   133200: contig of 8724 bp in length.
FEATURES             Location/Qualifiers
     source          1. .133200
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-18M17"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    38324 a  27723 c  27998 g  39103 t     52 others
ORIGIN


  Query Match           38.0%;  Score 418.2;  DB 43;  Length 133200;
  Best Local Similarity  77.2%;  Pred. No. 1.9e-110;
  Matches  521;  Conservative    0;  Mismatches  153;  Indels    1;  Gaps    1;

Qy    133 tgtctgctttgggggctgcagctccatacttgcaccatcccgctgacagtcacagtggcc 192
          ||| |||| |||||||  ||||||  || ||||| ||||| | ||  ||||||||||||
Db  99292 TGTTTGCTGTGGGGGCCACAGCTCTGTA-TTGCATCATCCTGGTGGTAGTCACAGTGGCC 99234

Qy    193 gggtcagtttcctgggttcccagccctctccagaagtgacggccgtggctcagctcttga 252
          | || |||||  ||||  ||||||   ||||||| ||| | ||  |||  |||||  |
Db  99233 GAGTGAGTTTTTTGGGGGCCCAGCTTCCTCCAGAGGTGGCAGCAATGGACCAGCTACTAG 99174

Qy    253 aggacttagacaggagcaccttcagaaagttgttgaaacttgtagtcggggccctgcatg 312
          |||||||  |||||||||  || |||||||||  ||||  |||| |||  |   ||||||  |
Db  99173 GGGACTTAAACAGGAGCATGTTTAGAAAGTTGCTGAAGTTTGTGGTCAGCAGCCTGCAAG 99114

Qy    313 ggaaagactgcagagaagctgtggagcaacttggtgccagcgccaacctgtcagaagagc 372
          ||  |  ||||||  |||| ||||||  ||   |||||  |  |||| ||||||||  |   |  |||
Db  99113 GGGAGGACTGCCGAGAGGCTGTGCAGTGTCTTGGGGTCAGCACCAACCTGCCGCAGGAGT 99054

Qy    373 gtctggccgtcctgctggcgggcacacacaccctgctccagcaggctctccggctgcccc 432
          | ||||  |  |||  |||||  ||||  |||||  ||||||||||||||  ||||  ||  ||||
Db  99053 GGCTGGGTGCCCTACTGGCAGGCATGCACACGCTGCTCCAGCAGGCCCTCCATCTTCCCC 98994

Qy    433 ctgctagtctaaagccagatgccttccaggaagagctccaggaacttggcattcctcagg 492
          |  |  ||  ||  |||||  |  |||||  |||  ||||||||||||||  ||||  ||  ||  |
Db  98993 CCACCAGCCTGAAGCCCAACACCTTCAGGGACCAGCTCCAGGAACTCTGCATCCCCCAAG 98934

Qy    493 atctaattggagatttggccagtttggcatttgggagtcaacgccctcttctcgactctg 552
          |  ||  |  ||  ||  ||||||||  |||  |||||||||  ||  ||  ||  ||  ||  ||||
Db  98933 ACCTGGTCGGGGACTTGGCCAGCGTGGTATTTGGGAGCCAGCGGCCCCTCCTTGATTCTG 98874

Qy    553 tagcccaacagcagggatcctcgctgcctcacgtgtcttacttccggtggcgggtggacg 612
          | |||||  ||||||||  |||  ||||||  ||  ||  ||  ||||  |  ||||  |||||||  |
Db  98873 TGGCCCAGCAGCAGGGGGCCTGGCTGCCCCATGTTGCTGACTTTCAGTGGTGGGTGGATG 98814

Qy    613 tggccatctcaaccagcgctcagtcccgctccctgcaaccgagtgttctcatgcagctga 672
          ||  |  |||||  |||||  ||  |  |  |  ||||  ||||||  |||||  ||  ||  ||||||||||
Db  98813 TGACAATCTCCACCAGTGCCCTGGCTCGCTTCCTGCAGCCGAGGGTCCTGATGCAGCTGA 98754

Qy    673 agctcacagatggatctgcacaccgcttcgaggtgcccatagccaaatttcaggagctgc 732
          ||||  |||||||  |  |||  ||||||||||||  ||||||||||  ||  ||||||||||
Db  98753 AGCTTTCAGATGGGTTAGCATACCGCTTCGAGGTCCCCATAGCCAAGTTCCAGGAGCTGC 98694
```

```
Qy    733 ggtacagtgtagccttggtccttaaggagatggcagaactggagaagaagtgtgagcgca 792
          ||||||| || ||| ||||||| |||||||||||||| |||||||||| ||||||| |||
Db  98693 GGTACAGCGTGGCCCTGGTCCTAAAGGAGATGGCAGATCTGGAGAAGAGGTGTGAGGGCA 98634

Qy    793 aactgcaggactgac 807
            |||||||||||||
Db  98633 GCCTGCAGGACTGAC 98619
```

```
RESULT   6
AF113540
LOCUS       AF113540     1588 bp    mRNA            PRI       04-DEC-1999
DEFINITION  Homo sapiens hypothalamus protein HT002 mRNA, complete cds.
ACCESSION   AF113540
VERSION     AF113540.1  GI:6523834
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1588)
  AUTHORS   Dong,H., Ren,S., Huang,C., Jiang,C., Li,Y., Zhou,J., Yu,Y., Xu,S.,
            Wang,Y., Fu,G., Chen,Z. and Han,Z.
  TITLE     A novel gene expressed in the human hypothalamus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1588)
  AUTHORS   Dong,H., Ren,S., Huang,C., Jiang,C., Li,Y., Zhou,J., Yu,Y., Xu,S.,
            Wang,Y., Fu,G., Chen,Z. and Han,Z.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-DEC-1998) Chinese National Human Genome Center at
            Shanghai, 351 Guo Shoujing Rd., Zhangjiang Hi-Tech Park, Pudong,
            Shanghai 201203, China
FEATURES             Location/Qualifiers
     source          1. .1588
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="hypothalamus"
     CDS             418. .1089
                     /codon_start=1
                     /product="hypothalamus protein HT002"
                     /protein_id="AAF14877.1"
                     /db_xref="GI:6523835"
                     /translation="MSAVGTATPYLHHPGDSHSGRVSFLGAQLPPEVAAMARLLGDLD
                     RSTFRKLLKFVVSSLQGEDCRDCCSVLGSAPTCRRAAGCLLAGMHTLLQQALRLPPTS
                     LKPDTFRDPAPGALHPPRPGRGLGQRGIWEPAALLDSVAQQQGAWLPHVADFRWRVDV
                     AISTSALARSLQPSVLMQLKLSDGSAYRFEVPTAKFQELRYSVALVLKEMADLEKRCE
                     RRLQD"
BASE COUNT      354 a    437 c    437 g    360 t
ORIGIN
```

```
  Query Match           37.5%;  Score 412.6;  DB 40;  Length 1588;
  Best Local Similarity   75.8%;  Pred. No. 5.6e-109;
  Matches  562;  Conservative    0;  Mismatches  174;  Indels    5;  Gaps    4;

Qy     84 tgtggttaggaccttctcttcggactggtcaagaaacgggaagaaaggatgtctgctttg 143
          | |||    ||||| ||   | | ||| |||| |    | || |   ||||||||| ||
Db    370 TCTGGGACCGACCTCCTGGGCTGGCTGATCAAAGAGGAAGCAGCAGCAATGTCTGCTGTG 429

Qy    144 ggggctgcagctccatacttgcaccatcccgctgacagtcacagtggccgggtcagtttc 203
          ||| ||||| |||||||| |||| ||||| | ||| |||||||||||||| || ||||||
Db    430 GGGACTGCAACTCCATACCTGCATCATCCTGGTGATAGTCACAGTGGCCGAGTGAGTTTC 489

Qy    204 ctgggttcccagccctctccagaagtgacggccgtggctcagctcttgaaggacttagac 263
          ||||  ||||||   ||||||| ||| | ||  |||| | |||  |   |||| |||||
Db    490 TTGGGGGCCCAGCTTCCTCCAGAGGTGGCAGCAATGGCCCGGCTACTAGGGGACCTAGAC 549

Qy    264 aggagcaccttcagaaagttgttgaaacttgtagtcggggccctgcatgggaaagactgc 323
          |||||||| |||||||||||| ||||  |||| ||| |   |||||| ||| | ||||||
Db    550 AGGAGCACGTTCAGAAAGTTGCTGAAGTTTGTGGTCAGCAGCCTGCAGGGGGAGGACTGC 609

Qy    324 agagaagctgtggagcaacttggtgccagcgccaacctgtcagaagagcgtctggccgtc 383
```

```
          ||||   | || |||  ||||| | |||||||||||| | ||     |   ||    |
Db    610 CGAGACTGT-TGCAGCGTCTTGGGGTCAGCGCCAACCTGCCGGAGAGCAGCTGGG--TGC 666

Qy    384 ctgctggcgggcacacacaccctgctccagcaggctctccggctgccccctgctagtcta 443
          |||||||| ||||  ||||| |||||||||||||| ||||| |||||||  | || ||
Db    667 CTGCTGGCAGGCATGCACACACTGCTCCAGCAGGCCCTCCGTCTGCCCCCCACCAGCCTG 726

Qy    444 aagccagatgccttccagga-agagctccaggaacttggcattcctcaggatctaattgg 502
          ||||| ||  |||||  |||   |||||||||| ||  |||| || || || ||  | ||
Db    727 AAGCCTGACACCTTCAGGGACCCAGCTCCAGGAGCTCTGCATCCCCCAAGACCTGGTCGG 786

Qy    503 agatttggccagtttggcatttgggagtcaacgccctcttctcgactctgtagcccaaca 562
          || ||||||||  ||| |||||||||| || || || || || || || ||||| ||||| ||
Db    787 GGACTTGGCCAGCGTGGTATTTGGGAGCCAGCGGCC-CTCCTTGATTCTGTGGCCCAGCA 845

Qy    563 gcagggatcctcgctgcctcacgtgtcttacttccggtggcgggtggacgtggccatctc 622
          ||||||  ||| |||||| || ||  || |||| |||||||||||||| || || |||||
Db    846 GCAGGGGGCCTGGCTGCCGCATGTTGCTGACTTTCGGTGGCGGGTGGATGTAGCAATCTC 905

Qy    623 aaccagcgctcagtcccgctccctgcaaccgagtgttctcatgcagctgaagctcacaga 682
          ||||| || | | | |||||||||||| ||||| || || ||||||||||||||  ||||
Db    906 CACCAGTGCCCTGGCTCGCTCCCTGCAGCCGAGCGTCCTGATGCAGCTGAAGCTTTCAGA 965

Qy    683 tggatctgcacaccgcttcgaggtgcccatagccaaatttcaggagctgcggtacagtgt 742
          ||| || ||| ||||||| ||||| |||| |||||| || ||||||||||||||||  ||
Db    966 TGGGTCAGCATACCGCTTTGAGGTCCCCACAGCCAAGTTCCAGGAGCTGCGGTACAGCGT 1025

Qy    743 agccttggtccttaaggagatggcagaactggagaagaagtgtgagcgcaaactgcagga 802
          ||| ||||||| |||||||||||||| |||||||||| |||||||||| |||||||||
Db   1026 GGCCCTGGTCCTAAAGGAGATGGCAGATCTGGAGAAGAGGTGTGAGCGCAGACTGCAGGA 1085

Qy    803 ctgactgaaccctggtactgt 823
          |||||    | || |   | ||
Db   1086 CTGACCCCTCACTTGACCAGT 1106
```

```
RESULT   7
AC011335
LOCUS       AC011335    78543 bp    DNA             HTG       19-DEC-1999
DEFINITION  Homo sapiens chromosome 5 clone CIT-HSPC_235N22, WORKING DRAFT
            SEQUENCE, 27 unordered pieces.
ACCESSION   AC011335
VERSION     AC011335.2  GI:6604456
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 78543)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 78543)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Dec 20, 1999 this sequence version replaced gi:6013605.
            -------Genome Center
            Center: Joint Genome Institute
            Web site: http://www.jgi.doe.gov
            -------Summary Statistics
            Estimated insert size: 78543; sum-of-contigs estimation
            Estimated insert size: 38000; pulse field gel estimation
            --------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 27 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
```

```
            *        1     1367: contig of 1367 bp in length
            *                    gap of unknown length
            *     1368     2437: contig of 1070 bp in length
            *                    gap of unknown length
            *     2438     3699: contig of 1262 bp in length
            *                    gap of unknown length
            *     3700     4899: contig of 1200 bp in length
            *                    gap of unknown length
            *     4900     6185: contig of 1286 bp in length
            *                    gap of unknown length
            *     6186     7215: contig of 1030 bp in length
            *                    gap of unknown length
            *     7216     8696: contig of 1481 bp in length
            *                    gap of unknown length
            *     8697    10413: contig of 1717 bp in length
            *                    gap of unknown length
            *    10414    11835: contig of 1422 bp in length
            *                    gap of unknown length
            *    11836    12914: contig of 1079 bp in length
            *                    gap of unknown length
            *    12915    14177: contig of 1263 bp in length
            *                    gap of unknown length
            *    14178    15971: contig of 1794 bp in length
            *                    gap of unknown length
            *    15972    17413: contig of 1442 bp in length
            *                    gap of unknown length
            *    17414    18482: contig of 1069 bp in length
            *                    gap of unknown length
            *    18483    19874: contig of 1392 bp in length
            *                    gap of unknown length
            *    19875    21384: contig of 1510 bp in length
            *                    gap of unknown length
            *    21385    23271: contig of 1887 bp in length
            *                    gap of unknown length
            *    23272    25113: contig of 1842 bp in length
            *                    gap of unknown length
            *    25114    26129: contig of 1016 bp in length
            *                    gap of unknown length
            *    26130    28129: contig of 2000 bp in length
            *                    gap of unknown length
            *    28130    30230: contig of 2101 bp in length
            *                    gap of unknown length
            *    30231    32849: contig of 2619 bp in length
            *                    gap of unknown length
            *    32850    35399: contig of 2550 bp in length
            *                    gap of unknown length
            *    35400    39211: contig of 3812 bp in length
            *                    gap of unknown length
            *    39212    41428: contig of 2217 bp in length
            *                    gap of unknown length
            *    41429    45655: contig of 4227 bp in length
            *                    gap of unknown length
            *    45656    78543: contig of 32888 bp in length.
FEATURES             Location/Qualifiers
     source          1. .78543
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CIT-HSPC_235N22"
BASE COUNT    24007 a  14700 c  15108 g  24701 t     27 others
ORIGIN


  Query Match             4.7%;  Score 51.6;  DB 44;  Length 78543;
  Best Local Similarity   59.6%;  Pred. No. 0.00059;
  Matches   87;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy     955 agctgggcagggtggcacagtttataatctcagcccttggaagtctgaggctggagaatg 1014
           |||  |||  |||||| ||   | ||||| |||| || || || | ||||| || | ||
Db   39290 AGCCAGGCGTGGTGGCTCACCCTTTAATCACAGCACTCGGGAGGCAGAGGCAGGCGGATT 39349

Qy    1015 ggaagtgtaagctgggcctggctttcatagtgaggctcagtgtcgaattaaagaggtaaa 1074
             | ||    ||||||  |  | || ||  | | || || ||  |   |   |||

Db   39350 TCTGAGTTCAGGCCAGCCTGGGCTATACAGAGAAACCCTGTCTCAAAAAACCAAAAAAAA 39409

Qy    1075 gcaactattaaaaaaaaaaaaaaaa 1100
             ||  |  |||||| |||||| |||
Db   39410 AAAAAAAAAAAAAAAGAAAAAAGAAA 39435


RESULT   8
AC007636
LOCUS       AC007636    78651 bp    DNA             ROD       15-FEB-2000
DEFINITION  Mus musculus chromosome 6 clone unknown strain RIII Fibroblast cell
            line C127, complete sequence.
ACCESSION   AC007636
VERSION     AC007636.19  GI:6862971
KEYWORDS    HTG.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 78651)
  AUTHORS   Yao,Z., Shual,S., Wu,H., Meisler,M. and Roe,B.A.
  TITLE     Mus musculus Chromosome 6 PAC Clone p1_4 In MND2 Region
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 78651)
  AUTHORS   Yao,Z., Shual,S., Wu,H., Meisler,M. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAY-1999) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 78651)
  AUTHORS   Yao,Z., Shual,S., Wu,H., Meisler,M. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-FEB-2000) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   4  (bases 1 to 78651)
  AUTHORS   Yao,Z., Shual,S., Wu,H., Meisler,M. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-FEB-2000) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Feb 3, 2000 this sequence version replaced gi:6730773.
FEATURES             Location/Qualifiers
     source          1. .78651
                     /organism="Mus musculus"
                     /strain="RIII Fibroblast cell line C127"
                     /db_xref="taxon:10090"
                     /chromosome="6"
                     /clone="unknown"
                     /clone_lib="Nat Sternberg mouse P1 RIII"
BASE COUNT    20921 a  20258 c  18946 g  18526 t
ORIGIN


  Query Match             4.6%;  Score 50.2;  DB 12;  Length 78651;
  Best Local Similarity   59.4%;  Pred. No. 0.0015;
  Matches   85;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy     955 agctgggcagggtggcacagtttataatctcagcccttggaagtctgaggctggagaatg 1014
           | ||||    |||||    |  || || | |||||  ||  || |||||   |||||||
Db   52591 AACTGGCTCTGGTGGAGAGGCCTAGAACCCCAGCCACTGAGAGACTGAGAGAGGAGAATC 52650

Qy    1015 ggaagtgtaagctgggcctggctttcatagtgaggctcagtgtcgaattaaagaggtaaa 1074
             ||||  |||    |||||    | || |||| |   || |||||   || |   ||
Db   52651 ACAAGTTCAAGGCCCACCTGGGCAACTTACTGAGACCTTGTCTCGAAAACAAAACAAAAC 52710

Qy    1075 gcaactattaaaaaaaaaaaaa 1097
             |||  | ||||| ||| ||||
Db   52711 AAAACATTAAAAAATAAATAAAA 52733


RESULT   9
AC007306
```

```
LOCUS       AC007306    78783 bp    DNA             ROD       15-FEB-2000
DEFINITION  Mus musculus chromosome 6 clone unknown strain RIII Fibroblast cell
            line C127, complete sequence.
ACCESSION   AC007306
VERSION     AC007306.15  GI:6958029
KEYWORDS    HTG.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 78783)
  AUTHORS   Chen,F., Do,T., Do,A., Meisler,M. and Roe,B.A.
  TITLE     Mus musculus Chromosome 6 P1 Clone p1_2 In MND2 Region
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 78783)
  AUTHORS   Chen,F., Do,T., Do,A., Meisler,M. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-APR-1999) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 78783)
  AUTHORS   Chen,F., Do,T., Do,A., Meisler,M. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-FEB-2000) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   4  (bases 1 to 78783)
  AUTHORS   Chen,F., Do,T., Do,A., Meisler,M. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-FEB-2000) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Feb 11, 2000 this sequence version replaced gi:6691340.
FEATURES             Location/Qualifiers
     source          1. .78783
                     /organism="Mus musculus"
                     /strain="RIII Fibroblast cell line C127"
                     /db_xref="taxon:10090"
                     /chromosome="6"
                     /clone="unknown"
                     /clone_lib="Nat Sternberg mouse P1 RIII"
BASE COUNT    20285 a  21224 c  19078 g  18196 t
ORIGIN


  Query Match             4.6%;  Score 50.2;  DB 12;  Length 78783;
  Best Local Similarity   59.4%;  Pred. No. 0.0015;
  Matches   85;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy     955 agctgggcagggtggcacagtttataatctcagcccttggaagtctgaggctggagaatg 1014
           | ||||    |||||    |  || || | |||||  ||  || |||||   |||||||
Db   20099 AACTGGCTCTGGTGGAGAGGCCTAGAACCCCAGCCACTGAGAGACTGAGAGAGGAGAATC 20158

Qy    1015 ggaagtgtaagctgggcctggctttcatagtgaggctcagtgtcgaattaaagaggtaaa 1074
             ||||  |||     |||||     | || |||| |   || |||||   || |   ||
Db   20159 ACAAGTTCAAGGCCCACCTGGGCAACTTACTGAGACCTTGTCTCGAAAACAAAACAAAAC 20218

Qy    1075 gcaactattaaaaaaaaaaaaa 1097
             |||  | |||||  ||| ||||
Db   20219 AAAACATTAAAAAATAAATAAAA 20241


RESULT  10
AC022781
LOCUS       AC022781   138608 bp    DNA             HTG       18-FEB-2000
DEFINITION  Mus musculus chromosome 11 clone RP23-199H17 map 11, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC022781
VERSION     AC022781.2  GI:6997225
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 138608)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 11, clone RP23-199H17
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 138608)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 18, 2000 this sequence version replaced gi:6922475.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5913
                Center clone name: 199_H_17
            --------------.
            * NOTE: This record contains 183 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      798: contig of 798 bp in length
            *                   gap of unknown length
            *      799     1617: contig of 819 bp in length
            *                   gap of unknown length
            *     1618     2414: contig of 797 bp in length
            *                   gap of unknown length
            *     2415     3240: contig of 826 bp in length
            *                   gap of unknown length
            *     3241     4035: contig of 795 bp in length
            *                   gap of unknown length
            *     4036     4857: contig of 822 bp in length
            *                   gap of unknown length
            *     4858     5686: contig of 829 bp in length
            *                   gap of unknown length
            *     5687     6490: contig of 804 bp in length
            *                   gap of unknown length
            *     6491     7300: contig of 810 bp in length
            *                   gap of unknown length
            *     7301     8124: contig of 824 bp in length
            *                   gap of unknown length
            *     8125     8943: contig of 819 bp in length
            *                   gap of unknown length
            *     8944     9746: contig of 803 bp in length
```

```
*                    gap of unknown length
*     9747    10557: contig of 811 bp in length
*                    gap of unknown length
*    10558    11370: contig of 813 bp in length
*                    gap of unknown length
*    11371    12200: contig of 830 bp in length
*                    gap of unknown length
*    12201    13037: contig of 837 bp in length
*                    gap of unknown length
*    13038    13864: contig of 827 bp in length
*                    gap of unknown length
*    13865    14685: contig of 821 bp in length
*                    gap of unknown length
*    14686    15499: contig of 814 bp in length
*                    gap of unknown length
*    15500    16295: contig of 796 bp in length
*                    gap of unknown length
*    16296    17111: contig of 816 bp in length
*                    gap of unknown length
*    17112    17916: contig of 805 bp in length
*                    gap of unknown length
*    17917    18720: contig of 804 bp in length
*                    gap of unknown length
*    18721    19540: contig of 820 bp in length
*                    gap of unknown length
*    19541    20336: contig of 796 bp in length
*                    gap of unknown length
*    20337    21137: contig of 801 bp in length
*                    gap of unknown length
*    21138    21952: contig of 815 bp in length
*                    gap of unknown length
*    21953    22748: contig of 796 bp in length
*                    gap of unknown length
*    22749    23524: contig of 776 bp in length
*                    gap of unknown length
*    23525    24339: contig of 815 bp in length
*                    gap of unknown length
*    24340    25157: contig of 818 bp in length
*                    gap of unknown length
*    25158    25951: contig of 794 bp in length
*                    gap of unknown length
*    25952    26750: contig of 799 bp in length
*                    gap of unknown length
*    26751    27560: contig of 810 bp in length
*                    gap of unknown length
*    27561    28362: contig of 802 bp in length
*                    gap of unknown length
*    28363    29174: contig of 812 bp in length
*                    gap of unknown length
*    29175    29970: contig of 796 bp in length
*                    gap of unknown length
*    29971    30780: contig of 810 bp in length
*                    gap of unknown length
*    30781    31583: contig of 803 bp in length
*                    gap of unknown length
*    31584    32387: contig of 804 bp in length
*                    gap of unknown length
*    32388    33191: contig of 804 bp in length
*                    gap of unknown length
*    33192    33988: contig of 797 bp in length
*                    gap of unknown length
*    33989    34790: contig of 802 bp in length
*                    gap of unknown length
*    34791    35568: contig of 778 bp in length
*                    gap of unknown length
*    35569    36389: contig of 821 bp in length
*                    gap of unknown length
*    36390    37184: contig of 795 bp in length
*                    gap of unknown length
*    37185    38006: contig of 822 bp in length
*                    gap of unknown length
*    38007    38813: contig of 807 bp in length
*                    gap of unknown length
*    38814    39651: contig of 838 bp in length
*                    gap of unknown length
*    39652    40433: contig of 782 bp in length
*                    gap of unknown length
*    40434    41248: contig of 815 bp in length
*                    gap of unknown length
*    41249    42072: contig of 824 bp in length
*                    gap of unknown length
*    42073    42878: contig of 806 bp in length
*                    gap of unknown length
*    42879    43683: contig of 805 bp in length
*                    gap of unknown length
*    43684    44504: contig of 821 bp in length
*                    gap of unknown length
*    44505    45326: contig of 822 bp in length
*                    gap of unknown length
*    45327    46144: contig of 818 bp in length
*                    gap of unknown length
*    46145    46952: contig of 808 bp in length
*                    gap of unknown length
*    46953    47795: contig of 843 bp in length
*                    gap of unknown length
*    47796    48618: contig of 823 bp in length
*                    gap of unknown length
*    48619    49411: contig of 793 bp in length
*                    gap of unknown length
*    49412    50228: contig of 817 bp in length
*                    gap of unknown length
*    50229    51055: contig of 827 bp in length
*                    gap of unknown length
*    51056    51871: contig of 816 bp in length
*                    gap of unknown length
*    51872    52678: contig of 807 bp in length
*                    gap of unknown length
*    52679    53493: contig of 815 bp in length
*                    gap of unknown length
*    53494    54303: contig of 810 bp in length
*                    gap of unknown length
*    54304    55085: contig of 782 bp in length
*                    gap of unknown length
*    55086    55921: contig of 836 bp in length
*                    gap of unknown length
*    55922    56688: contig of 767 bp in length
*                    gap of unknown length
*    56689    57505: contig of 817 bp in length

  Query Match             4.4%;  Score 48.6;  DB 56;  Length 138608;
  Best Local Similarity   60.0%;  Pred. No. 0.0046;
  Matches   81;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy    954 cagctgggcagggtggcacagtttataatctcagcccttggaagtctgaggctggagaat 1013
          | || ||||| |||||||||    | ||||  |||| ||    || ||||||| |||||||
Db  64306 CTGCCGGGCATGGTGGCACATACTGTAATTCCAGCACTCAAGAGACTGAGGCAGGAGAAT 64365

Qy   1014 gggaagtgtaagctgggcctggctttcatagtgaggctcagtgtcgaattaaagaggtaa 1073
          || | | ||    |||||    |||| |||| || |    ||  || | ||
Db  64366 CACAAATTTGAGACTAGCCTGGGCCACATAATGAGATTCTATTAAAAAAATAATAAATAT 64425

Qy   1074 agcaactattaaaaa 1088
          |   |    |||| |
Db  64426 AATGAAACCTAAATA 64440


RESULT  11
AC016365
LOCUS       AC016365   167859 bp    DNA             HTG       17-FEB-2000
DEFINITION  Mus musculus clone RP23-450A18, WORKING DRAFT SEQUENCE, 30
            unordered pieces.
ACCESSION   AC016365
VERSION     AC016365.3  GI:6984431
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 167859)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-450A18
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 167859)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 17, 2000 this sequence version replaced gi:6730840.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5030
                Center clone name: 450_A_18
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 152833 bases at least Q40
                Consensus quality: 161208 bases at least Q30
                Consensus quality: 164488 bases at least Q20
                Insert size: 181000; agarose-fp
                Insert size: 167859; sum-of-contigs
                Quality coverage: 3.1 in Q20 bases; agarose-fp
                Quality coverage: 3.3 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 30 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1125: contig of 1125 bp in length
            *                   gap of unknown length
            *     1126     2198: contig of 1073 bp in length
            *                   gap of unknown length
            *     2199     3443: contig of 1245 bp in length
            *                   gap of unknown length
            *     3444     4856: contig of 1413 bp in length
            *                   gap of unknown length
            *     4857     7532: contig of 2676 bp in length
            *                   gap of unknown length
            *     7533     9604: contig of 2072 bp in length
            *                   gap of unknown length
            *     9605    12241: contig of 2637 bp in length
            *                   gap of unknown length
            *    12242    14823: contig of 2582 bp in length
            *                   gap of unknown length
            *    14824    16854: contig of 2031 bp in length
            *                   gap of unknown length
            *    16855    19484: contig of 2630 bp in length
            *                   gap of unknown length
            *    19485    21373: contig of 1889 bp in length
            *                   gap of unknown length
            *    21374    23406: contig of 2033 bp in length
            *                   gap of unknown length
            *    23407    26007: contig of 2601 bp in length
            *                   gap of unknown length
            *    26008    29060: contig of 3053 bp in length
            *                   gap of unknown length
            *    29061    31821: contig of 2761 bp in length
            *                   gap of unknown length
            *    31822    37887: contig of 6066 bp in length
            *                   gap of unknown length
            *    37888    41490: contig of 3603 bp in length
            *                   gap of unknown length
            *    41491    47656: contig of 6166 bp in length
            *                   gap of unknown length
            *    47657    55135: contig of 7479 bp in length
            *                   gap of unknown length
            *    55136    60708: contig of 5573 bp in length
            *                   gap of unknown length
            *    60709    68423: contig of 7715 bp in length
            *                   gap of unknown length
            *    68424    73176: contig of 4753 bp in length
            *                   gap of unknown length
            *    73177    82052: contig of 8876 bp in length
            *                   gap of unknown length
            *    82053    90689: contig of 8637 bp in length
            *                   gap of unknown length
            *    90690    98893: contig of 8204 bp in length
            *                   gap of unknown length
            *    98894   109482: contig of 10589 bp in length
            *                   gap of unknown length
            *   109483   120462: contig of 10980 bp in length
            *                   gap of unknown length
            *   120463   131928: contig of 11466 bp in length
            *                   gap of unknown length
            *   131929   148037: contig of 16109 bp in length
            *                   gap of unknown length
            *   148038   167859: contig of 19822 bp in length.
FEATURES             Location/Qualifiers
     source          1. .167859
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP23-450A18"
                     /clone_lib="RPCI-23 Female Mouse BAC"
BASE COUNT    48301 a  36172 c  36115 g  47258 t     13 others
ORIGIN


  Query Match             4.4%;  Score 48.4;  DB 55;  Length 167859;
  Best Local Similarity   64.9%;  Pred. No. 0.0053;
  Matches   87;  Conservative    0;  Mismatches   46;  Indels    1;  Gaps    1;

Qy    941 ttttatgagtcaacagctgggcagggtggcacag-tttataatctcagcccttggaagtc 999
          ||| | ||| |   | |||||   | ||||||||   |  ||||||||| |||||| | |
Db 117520 TTTCAAGAGACCTAAACTGGGTGTGATGGCACAGACCTGCAATCTCAGCACTTGGATGGC 117579

Qy   1000 tgaggctggagaatgggaagtgtaagctgggcctggctttcatagtgaggctcagtgtcg 1059
          ||||   | || ||  |  ||  |||    |||||| ||  |||||||| |  ||| ||
Db 117580 TGAGTGGGAAGGATCAGGGGTTCAAGGCCAGCCTGGGTTATATAGTGAGACCTAGTCTCA 117639

Qy   1060 aattaaagaggtaa 1073
          ||   || | ||||
Db 117640 AAAACAAAAAGTAA 117653


RESULT  12
AC015891
LOCUS       AC015891   230475 bp    DNA             HTG       17-NOV-1999.
DEFINITION  Mus musculus chromosome 10 clone RP21-522L13 map 10, LOW-PASS
```

```
            SEQUENCE SAMPLING.
ACCESSION   AC015891
VERSION     AC015891.1  GI:6446777
KEYWORDS    HTG; HTGS_PHASE0; NULL.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 230475)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 10, clone RP21-522L13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 230475)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Project Information
                Center project name: L761
                Center clone name: 522_L_13
            --------------.
            * NOTE: This record contains 279 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      827: contig of 827 bp in length
            *                    gap of unknown length
            *      828     1737: contig of 910 bp in length
            *                    gap of unknown length
            *     1738     2655: contig of 918 bp in length
            *                    gap of unknown length
            *     2656     3510: contig of 855 bp in length
            *                    gap of unknown length
            *     3511     4384: contig of 874 bp in length
            *                    gap of unknown length
            *     4385     5298: contig of 914 bp in length
            *                    gap of unknown length
            *     5299     6231: contig of 933 bp in length
            *                    gap of unknown length
            *     6232     7060: contig of 829 bp in length
            *                    gap of unknown length
            *     7061     7952: contig of 892 bp in length
            *                    gap of unknown length
            *     7953     8847: contig of 895 bp in length
            *                    gap of unknown length
            *     8848     9706: contig of 859 bp in length
            *                    gap of unknown length
            *     9707    10550: contig of 844 bp in length
            *                    gap of unknown length
            *    10551    11431: contig of 881 bp in length
            *                    gap of unknown length
            *    11432    12384: contig of 953 bp in length
            *                    gap of unknown length
            *    12385    13248: contig of 864 bp in length
            *                    gap of unknown length
            *    13249    14083: contig of 835 bp in length
            *                    gap of unknown length
            *    14084    14955: contig of 872 bp in length
            *                    gap of unknown length
            *    14956    15823: contig of 868 bp in length
            *                    gap of unknown length
            *    15824    16675: contig of 852 bp in length
            *                    gap of unknown length
            *    16676    17644: contig of 969 bp in length
            *                    gap of unknown length
            *    17645    18541: contig of 897 bp in length
            *                    gap of unknown length
            *    18542    19391: contig of 850 bp in length
            *                    gap of unknown length
            *    19392    20239: contig of 848 bp in length
            *                    gap of unknown length
            *    20240    21078: contig of 839 bp in length
            *                    gap of unknown length
            *    21079    21897: contig of 819 bp in length
            *                    gap of unknown length
            *    21898    22812: contig of 915 bp in length
            *                    gap of unknown length
            *    22813    23708: contig of 896 bp in length
            *                    gap of unknown length
            *    23709    24563: contig of 855 bp in length
            *                    gap of unknown length
            *    24564    25430: contig of 867 bp in length
            *                    gap of unknown length
            *    25431    26328: contig of 898 bp in length
            *                    gap of unknown length
            *    26329    27131: contig of 803 bp in length
            *                    gap of unknown length
            *    27132    28001: contig of 870 bp in length
            *                    gap of unknown length
            *    28002    28961: contig of 960 bp in length
            *                    gap of unknown length
            *    28962    29913: contig of 952 bp in length
            *                    gap of unknown length
            *    29914    30776: contig of 863 bp in length
            *                    gap of unknown length
            *    30777    31632: contig of 856 bp in length
            *                    gap of unknown length
            *    31633    32525: contig of 893 bp in length
            *                    gap of unknown length
            *    32526    33367: contig of 842 bp in length
            *                    gap of unknown length
            *    33368    34288: contig of 921 bp in length
            *                    gap of unknown length
            *    34289    35097: contig of 809 bp in length
            *                    gap of unknown length
            *    35098    36050: contig of 953 bp in length
            *                    gap of unknown length
            *    36051    36942: contig of 892 bp in length
            *                    gap of unknown length
            *    36943    37783: contig of 841 bp in length
            *                    gap of unknown length
            *    37784    38640: contig of 857 bp in length
            *                    gap of unknown length
            *    38641    39503: contig of 863 bp in length
            *                    gap of unknown length
            *    39504    40441: contig of 938 bp in length
            *                    gap of unknown length
            *    40442    41319: contig of 878 bp in length
            *                    gap of unknown length
            *    41320    42139: contig of 820 bp in length
            *                    gap of unknown length
            *    42140    42883: contig of 744 bp in length
            *                    gap of unknown length
```

```
            *    42884    43606: contig of 723 bp in length
            *                    gap of unknown length
            *    43607    44330: contig of 724 bp in length
            *                    gap of unknown length
            *    44331    45057: contig of 727 bp in length
            *                    gap of unknown length
            *    45058    45787: contig of 730 bp in length
            *                    gap of unknown length
            *    45788    46513: contig of 726 bp in length
            *                    gap of unknown length
            *    46514    47241: contig of 728 bp in length
            *                    gap of unknown length
            *    47242    47965: contig of 724 bp in length
            *                    gap of unknown length
            *    47966    48688: contig of 723 bp in length
            *                    gap of unknown length
            *    48689    49426: contig of 738 bp in length
            *                    gap of unknown length
            *    49427    50142: contig of 716 bp in length
            *                    gap of unknown length
            *    50143    50889: contig of 747 bp in length
            *                    gap of unknown length
            *    50890    51613: contig of 724 bp in length
            *                    gap of unknown length
            *    51614    52344: contig of 731 bp in length
            *                    gap of unknown length
            *    52345    53080: contig of 736 bp in length
            *                    gap of unknown length
            *    53081    53840: contig of 760 bp in length
            *                    gap of unknown length
            *    53841    54571: contig of 731 bp in length
            *                    gap of unknown length
            *    54572    55300: contig of 729 bp in length
            *                    gap of unknown length
            *    55301    56024: contig of 724 bp in length
            *                    gap of unknown length
            *    56025    56765: contig of 741 bp in length
            *                    gap of unknown length
            *    56766    57489: contig of 724 bp in length
            *                    gap of unknown length
            *    57490    58214: contig of 725 bp in length
            *                    gap of unknown length
            *    58215    58937: contig of 723 bp in length
            *                    gap of unknown length
            *    58938    59658: contig of 721 bp in length
            *                    gap of unknown length
            *    59659    60380: contig of 722 bp in length
            *                    gap of unknown length
            *    60381    61105: contig of 725 bp in length
            *                    gap of unknown length
            *    61106    61830: contig of 725 bp in length
            *                    gap of unknown length

  Query Match             4.3%;  Score 47.6;  DB 42;  Length 230475;
  Best Local Similarity   61.8%;  Pred. No. 0.0093;
  Matches   94;  Conservative    0;  Mismatches   54;  Indels    4;  Gaps    1;

Qy     952 aacagctgggcagggtggcacagtttataatctcagcccttggaagtctgaggctggaga 1011
           | | |||||||  ||||||  ||   | |||||  |||| |||||  ||  |  ||||  ||  ||
Db  100621 ACCTGCTGGGCGTGGTGGCGCACGCTTTAATCCCAGCACTTGGGAGGCAAAGGCAGGTGA 100680

Qy    1012 atgggaagtgtaagctgggcctggctttcatagtgaggctcagtgtcga----attaaag 1067
             |    |||  |  ||    ||||||   |  ||   |||||    |||     |      | || |
Db  100681 TTTCTGAGTTTGAGGCCAGCCTGGTCTACAGTGTGAGTTCCAGGACAGCCAGGACTACAC 100740

Qy    1068 aggtaaagcaactattaaaaaaaaaaaaaaaa 1099
           |||  |||  |    |  ||  |||||  ||||  ||  |
Db  100741 AGGAAAACCCTATCTTGAAAAACAAAACAACA 100772


RESULT  13
AC020965/c
LOCUS       AC020965   238554 bp    DNA             HTG       10-FEB-2000
DEFINITION  Mus musculus clone RP21-82I16, WORKING DRAFT SEQUENCE, 80 unordered
            pieces.
ACCESSION   AC020965
VERSION     AC020965.1  GI:6691265
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 238554)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Mouse
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 238554)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------Summary Statistics
            Consensus quality: 160205 bases at least Q40
            Consensus quality: 219616 bases at least Q30
            Consensus quality: 238602 bases at least Q20
            Estimated insert size: 238554; sum-of-contigs estimation
            Estimated insert size: 162000; pulse field gel estimation
            Quality coverage:  4.94x in Q20 bases; pulse field gel estimation
            Quality coverage:  3.36x in Q20 bases; sum-of-contigs estimation
            -------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 80 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1253: contig of 1253 bp in length
            *                    gap of unknown length
            *     1254     2582: contig of 1329 bp in length
            *                    gap of unknown length
            *     2583     3769: contig of 1187 bp in length
            *                    gap of unknown length
            *     3770     4974: contig of 1205 bp in length
            *                    gap of unknown length
            *     4975     6214: contig of 1240 bp in length
            *                    gap of unknown length
            *     6215     7484: contig of 1270 bp in length
            *                    gap of unknown length
            *     7485     8808: contig of 1324 bp in length
            *                    gap of unknown length
            *     8809    10079: contig of 1271 bp in length
            *                    gap of unknown length
            *    10080    11473: contig of 1394 bp in length
            *                    gap of unknown length
            *    11474    12921: contig of 1448 bp in length
            *                    gap of unknown length
            *    12922    14128: contig of 1207 bp in length
            *                    gap of unknown length
            *    14129    15531: contig of 1403 bp in length
            *                    gap of unknown length
            *    15532    17044: contig of 1513 bp in length
            *                    gap of unknown length
            *    17045    18665: contig of 1621 bp in length
            *                    gap of unknown length
            *    18666    19891: contig of 1226 bp in length
            *                    gap of unknown length
            *    19892    21416: contig of 1525 bp in length
            *                    gap of unknown length
            *    21417    22724: contig of 1308 bp in length
            *                    gap of unknown length
```

```
*    22725    24061: contig of 1337 bp in length
*                    gap of unknown length
*    24062    25240: contig of 1179 bp in length
*                    gap of unknown length
*    25241    26539: contig of 1299 bp in length
*                    gap of unknown length
*    26540    27990: contig of 1451 bp in length
*                    gap of unknown length
*    27991    29324: contig of 1334 bp in length
*                    gap of unknown length
*    29325    30605: contig of 1281 bp in length
*                    gap of unknown length
*    30606    31933: contig of 1328 bp in length
*                    gap of unknown length
*    31934    33488: contig of 1555 bp in length
*                    gap of unknown length
*    33489    34824: contig of 1336 bp in length
*                    gap of unknown length
*    34825    36237: contig of 1413 bp in length
*                    gap of unknown length
*    36238    37659: contig of 1422 bp in length
*                    gap of unknown length
*    37660    38946: contig of 1287 bp in length
*                    gap of unknown length
*    38947    40159: contig of 1213 bp in length
*                    gap of unknown length
*    40160    41958: contig of 1799 bp in length
*                    gap of unknown length
*    41959    43508: contig of 1550 bp in length
*                    gap of unknown length
*    43509    45153: contig of 1645 bp in length
*                    gap of unknown length
*    45154    46735: contig of 1582 bp in length
*                    gap of unknown length
*    46736    48167: contig of 1432 bp in length
*                    gap of unknown length
*    48168    50115: contig of 1948 bp in length
*                    gap of unknown length
*    50116    52053: contig of 1938 bp in length
*                    gap of unknown length
*    52054    53298: contig of 1245 bp in length
*                    gap of unknown length
*    53299    55095: contig of 1797 bp in length
*                    gap of unknown length
*    55096    56476: contig of 1381 bp in length
*                    gap of unknown length
*    56477    57826: contig of 1350 bp in length
*                    gap of unknown length
*    57827    59160: contig of 1334 bp in length
*                    gap of unknown length
*    59161    60898: contig of 1738 bp in length
*                    gap of unknown length
*    60899    62119: contig of 1221 bp in length
*                    gap of unknown length
*    62120    63994: contig of 1875 bp in length
*                    gap of unknown length
*    63995    65567: contig of 1573 bp in length
*                    gap of unknown length
*    65568    66865: contig of 1298 bp in length
*                    gap of unknown length
*    66866    68057: contig of 1192 bp in length
*                    gap of unknown length
*    68058    69774: contig of 1717 bp in length
*                    gap of unknown length
*    69775    72153: contig of 2379 bp in length
*                    gap of unknown length
*    72154    75289: contig of 3136 bp in length
*                    gap of unknown length
*    75290    78389: contig of 3100 bp in length
*                    gap of unknown length
*    78390    80910: contig of 2521 bp in length
*                    gap of unknown length
*    80911    82153: contig of 1243 bp in length
*                    gap of unknown length
*    82154    84007: contig of 1854 bp in length
*                    gap of unknown length
*    84008    86443: contig of 2436 bp in length
*                    gap of unknown length
*    86444    89640: contig of 3197 bp in length
*                    gap of unknown length
*    89641    91024: contig of 1384 bp in length
*                    gap of unknown length
*    91025    93347: contig of 2323 bp in length
*                    gap of unknown length
*    93348    95487: contig of 2140 bp in length
*                    gap of unknown length
*    95488    98489: contig of 3002 bp in length
*                    gap of unknown length
*    98490   100499: contig of 2010 bp in length
*                    gap of unknown length
*   100500   104757: contig of 4258 bp in length
*                    gap of unknown length
*   104758   106505: contig of 1748 bp in length
*                    gap of unknown length
*   106506   112258: contig of 5753 bp in length
*                    gap of unknown length
*   112259   115111: contig of 2853 bp in length
*                    gap of unknown length
*   115112   117906: contig of 2795 bp in length
*                    gap of unknown length
*   117907   120203: contig of 2297 bp in length
*                    gap of unknown length
*   120204   125155: contig of 4952 bp in length
*                    gap of unknown length
*   125156   132035: contig of 6880 bp in length
*                    gap of unknown length
*   132036   138866: contig of 6831 bp in length
*                    gap of unknown length
*   138867   146355: contig of 7489 bp in length
*                    gap of unknown length
*   146356   155442: contig of 9087 bp in length
*                    gap of unknown length
*   155443   161962: contig of 6520 bp in length
*                    gap of unknown length
*   161963   170957: contig of 8995 bp in length
*                    gap of unknown length
*   170958   181490: contig of 10533 bp in length
*                    gap of unknown length
*   181491   193728: contig of 12238 bp in length
*                    gap of unknown length
*   193729   204122: contig of 10394 bp in length
*                    gap of unknown length
*   204123   223034: contig of 18912 bp in length
*                    gap of unknown length
*   223035   238554: contig of 15520 bp in length.
FEATURES             Location/Qualifiers

  Query Match          4.2%;  Score 46.6;  DB 55;  Length 238554;
  Best Local Similarity   61.7%;  Pred. No. 0.018;
  Matches   92;  Conservative    0;  Mismatches   54;  Indels    3;  Gaps    1;

Qy     955 agctgggcagggtggcacagtttataatctcagcccttggaagtctgaggctggagaatg 1014
           ||||||| ||  |         | ||||| |||| || || || | ||||| || | ||
Db   88412 AGCTGGGTGGGTGGTGCACACCTTTAATCCCAGCACTCGGGAGGCAGAGGCAGGCGGATT 88353

Qy    1015 ggaagtgtaagctgggcctggctttcatagtgaggctcagt---gtcgaattaaagaggt 1071
            |||  |||    ||||||  | || ||||||   |||    | ||   || | ||
Db   88352 CTGAGTTCAAGGCCAGCCTGGTCTACAAAGTGAGTTCCAGAACAGCCGGGCTATACAGAG 88293

Qy    1072 aaagcaactattaaaaaaaaaaaaaaaa 1100
           ||| |   | | ||||||||||||||||
Db   88292 AAACCCTGTCTCAAAAAAAAAAAAAAAA 88264


RESULT  14
AF213391
```

```
LOCUS       AF213391      963 bp    mRNA            ROD       07-FEB-2000
DEFINITION  Mus musculus ATP-binding cassette protein (Abcb8) mRNA, partial
            cds.
ACCESSION   AF213391
VERSION     AF213391.1  GI:6911206
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 963)
  AUTHORS   Schriml,L.M. and Dean,M.
  TITLE     Identification of 18 mouse ABC genes and characterization of the
            ABC superfamily in Mus musculus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 963)
  AUTHORS   Schriml,L.M. and Dean,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Laboratory of Genomic Diversity, National
            Cancer Institute, FCRDC, P.O.Box B, Frederick, MD 21702-1201, USA
FEATURES             Location/Qualifiers
     source          1. .963
                     /organism="Mus musculus"
                     /strain="C57BL/6JxDBA"
                     /db_xref="taxon:10090"
                     /note="sub-family B; MDR member 8"
     gene            <1. .963
                     /gene="Abcb8"
     CDS             <1. .394
                     /gene="Abcb8"
                     /codon_start=2
                     /product="ATP-binding cassette protein"
                     /protein_id="AAF31430.1"
                     /db_xref="GI:6911207"
                     /translation="QTWLQCAQGLHPEAALWKIVALVGQSGGGKTTVASLLERFYDPE
                     AGSVTLDGHDLRTLNPSWLRGQVIGFISQEPVLFATTIMENIRFGKLDASDEEVYTAA
                     REANAHEFISSFHDGYSTVVGWDPRRTP"
BASE COUNT      242 a    266 c    263 g    192 t
ORIGIN
```

```
  Query Match             4.2%;  Score 45.8;  DB 12;  Length 963;
  Best Local Similarity   57.2%;  Pred. No. 0.019;
  Matches   83;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy     956 gctgggcagggtggcacagtttataatctcagcccttggaagtctgaggctggagaatgg 1015
           ||  |||  ||| | |  | || ||||| |||| ||||| || | || |  ||||  |
Db     816 GCCAGGCCTGGTAGAATGGGTTGTAATCCCAGCACTTGGGAGACAGAAGTGGGAGGGTCT 875

Qy    1016 gaagtgtaagctgggcctggctttcatagtgaggctcagtgtcgaattaaagaggtaaag 1075
           | |  |||     |||||  |    ||     |||       ||  ||| |  | ||
Db     876 AGAATTCAAGGCCATCCTGGGCTAGTGAGACCCTCTCTCAACAAAAAGAAAAACCTCAAT 935

Qy    1076 caactattaaaaaaaaaaaaaaaa 1100
            ||  |  ||||||||||||||||||
Db     936 TAAAAAAAAAAAAAAAAAAAAAAA 960
```

```
RESULT  15
I66494/c
LOCUS       I66494       7218 bp    DNA             PAT       28-DEC-1997
DEFINITION  Sequence 14 from patent US 5670367.
ACCESSION   I66494
VERSION     I66494.1  GI:2724471
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7218)
  AUTHORS   Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
  TITLE     Recombinant fowlpox virus
  JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
FEATURES             Location/Qualifiers
     source          1. .7218
                     /organism="unknown"
BASE COUNT     1944 a    1491 c    1486 g    1929 t     368 others
ORIGIN
```

```
  Query Match             4.1%;  Score 45.6;  DB 5;  Length 7218;
  Best Local Similarity   1.7%;  Pred. No. 0.026;
  Matches    6;  Conservative  207;  Mismatches  141;  Indels    0;  Gaps    0;

Qy     747 ttggtccttaaggagatggcaggaactggagaagaagtgtgagcgcaaactgcaggactga 806
           |||||  |  ::::::: :: ::::  :::::::::: : ::: : :::  : ::::  ::
Db    1441 TTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1382

Qy     807 ctgaaccctggtactgtgggtgctgaagctggtaccagaacacagccccccactggtgat 866
             :::    :: :  : ::: :  ::::  :: :  :::: : ::       :  :: ::
Db    1381 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1322

Qy     867 gagcccaactccattgaggtcctgcatgtgagaacgtattttaagtgaaaagacagcggg 926
           :::   ::    :  ::::   : : : ::::: : :    ::: ::::::: :: :::
Db    1321 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1262

Qy     927 actttcaggttttgttttatgagtcaacagctgggcagggtggcacagtttataatctca 986
           :     :::    :    : :::  :: ::  ::: :::: :: : ::   : ::    :
Db    1261 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1202

Qy     987 gcccttggaagtctgaggctggagaatgggaagtgtaagctgggcctggctttcatagtg 1046
           :     :::::   ::::  :::::: :::::: : :::  :::   ::     : :: :
Db    1201 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1142

Qy    1047 aggctcagtgtcgaattaaagaggtaaagcaactattaaaaaaaaaaaaaaaa 1100
           :::   :: :  :::  ::::::: :::: ::  :  ::::::::::::::::
Db    1141 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1088
```

```
Search completed: May 15, 2000, 01:54:21
Job time: 11180 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: May 15, 2000, 01:54:21 ; Search time 1646.42 Seconds
(without alignments)
-800.605 Million cell updates/sec

Title: US-09-223-796-3
Perfect score: 1355
Sequence: 1 gggcaggcagttgaggtgga.....gtgtttcaggcagggcccgg 1355

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vi:*
17: em_fun:*
18: em_hum1:*
19: em_hum2:*
20: em_in:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pl:*
27: em_ro:*
28: em_sts:*
29: em_sy:*
30: em_un:*
31: em_vi:*
32: gb_htg1:*
33: gb_htg2:*
34: gb_in1:*
35: gb_in2:*
36: em_ba1:*
37: em_ba2:*
38: em_hum3:*
39: em_hum4:*
40: gb_pr4:*
41: gb_htg3:*
42: gb_htg4:*
43: gb_htg5:*
44: gb_htg6:*
45: gb_htg7:*
46: em_htg1:*
47: em_htg2:*
48: em_htg3:*
49: em_hum5:*
50: gb_pl3:*
51: gb_pr5:*
52: gb_htg8:*
53: gb_htg9:*
54: gb_htg10:*
55: gb_htg11:*
56: gb_htg12:*
57: gb_htg13:*
58: gb_htg14:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1144.4	84.5	1588	40	AF113540	AF113540 Homo sapi
c	2	1085.8	80.1	108369	40	AF124523	AF124523 Homo sapi
c	3	1084.2	80.0	174923	52	AF216665	AF216665 Homo sapi
	4	1081	79.8	128009	40	AF118808	AF118808 Homo sapi
	5	1077.8	79.5	69528	40	AF146367	AF146367 Homo sapi
c	6	724.2	53.4	133200	43	AC011951	AC011951 Homo sapi
	7	110.8	8.2	347	9	HS79C3R	Z63154 H.sapiens C
	8	104	7.7	1041	40	AF014882	AF014882 Homo sapi
	9	104	7.7	1041	40	AF014883	AF014883 Homo sapi
	10	104	7.7	1041	40	AF014884	AF014884 Homo sapi
	11	104	7.7	1041	40	AF014885	AF014885 Homo sapi
	12	104	7.7	1041	40	AF014886	AF014886 Homo sapi
	13	104	7.7	1041	40	AF014887	AF014887 Homo sapi
	14	104	7.7	1041	40	AF014888	AF014888 Homo sapi
	15	104	7.7	1041	40	AF014889	AF014889 Homo sapi
	16	104	7.7	1041	40	AF014890	AF014890 Homo sapi
	17	104	7.7	1041	40	AF014891	AF014891 Homo sapi
	18	104	7.7	1041	40	AF014892	AF014892 Homo sapi
	19	104	7.7	1041	40	AF014893	AF014893 Homo sapi
	20	104	7.7	1041	40	AF014894	AF014894 Homo sapi
	21	104	7.7	1041	40	AF014895	AF014895 Homo sapi
	22	104	7.7	1041	40	AF014896	AF014896 Homo sapi
	23	104	7.7	1041	40	AF014897	AF014897 Homo sapi
	24	104	7.7	1041	40	AF014898	AF014898 Homo sapi
	25	104	7.7	1041	40	AF014899	AF014899 Homo sapi
	26	104	7.7	1041	40	AF014900	AF014900 Homo sapi
	27	104	7.7	15447	10	MIHSGENOM	X62996 H.sapiens m
	28	104	7.7	16559	9	HUMMTA	D38112 Homo sapien
	29	104	7.7	16570	11	HSMITG	X93334 Homo sapien
	30	102.4	7.6	1041	40	AF014901	AF014901 Homo sapi
c	31	102.4	7.6	1320	40	U95646	U95646 Homo sapien
	32	102.4	7.6	2771	10	HUMMTM1	M10546 Human mitoc
	33	102.4	7.6	16569	10	MIHSXX	V00662 H.sapiens m
	34	102.4	7.6	16569	51	HUMMTCG	J01415 Human mitoc
	35	101.8	7.5	5840	40	AF134583	AF134583 Homo sapi
	36	91.2	6.7	1041	11	AF014903	AF014903 Pan trogl
	37	91.2	6.7	16563	9	CHPMTE	D38116 Pan paniscu
	38	89.6	6.6	1041	11	AF014904	AF014904 Pan trogl
	39	89	6.6	1041	11	AF014902	AF014902 Pan trogl
	40	89	6.6	1041	11	AF014905	AF014905 Pan trogl
	41	89	6.6	1041	11	AF014907	AF014907 Pan trogl
	42	89	6.6	1041	11	AF014911	AF014911 Pan trogl
	43	89	6.6	1041	11	AF014915	AF014915 Pan trogl
	44	89	6.6	1041	11	AF014919	AF014919 Pan trogl
	45	89	6.6	1041	11	AF014921	AF014921 Pan trogl

ALIGNMENTS

```
RESULT   1
AF113540
LOCUS       AF113540     1588 bp    mRNA            PRI       04-DEC-1999
DEFINITION  Homo sapiens hypothalamus protein HT002 mRNA, complete cds.
ACCESSION   AF113540
VERSION     AF113540.1  GI:6523834
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1588)
  AUTHORS   Dong,H., Ren,S., Huang,C., Jiang,C., Li,Y., Zhou,J., Yu,Y., Xu,S.,
            Wang,Y., Fu,G., Chen,Z. and Han,Z.
  TITLE     A novel gene expressed in the human hypothalamus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1588)
  AUTHORS   Dong,H., Ren,S., Huang,C., Jiang,C., Li,Y., Zhou,J., Yu,Y., Xu,S.,
            Wang,Y., Fu,G., Chen,Z. and Han,Z.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-DEC-1998) Chinese National Human Genome Center at
            Shanghai, 351 Guo Shoujing Rd., Zhangjiang Hi-Tech Park, Pudong,
            Shanghai 201203, China
FEATURES             Location/Qualifiers
     source          1. .1588
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="hypothalamus"
     CDS             418. .1089
                     /codon_start=1
                     /product="hypothalamus protein HT002"
                     /protein_id="AAF14877.1"
                     /db_xref="GI:6523835"
                     /translation="MSAVGTATPYLHHPGDSHSGRVSFLGAQLPPEVAAMARLLGDLD
                     RSTFRKLLKFVVSSLQGEDCRDCCSVLGSAPTCRRAAGCLLAGMHTLLQQALRLPPTS
                     LKPDTFRDPAPGALHPPRPGRGLGQRGIWEPAALLDSVAQQQGAWLPHVADFRWRVDV
                     AISTSALARSLQPSVLMQLKLSDGSAYRFEVPTAKFQELRYSVALVLKEMADLEKRCE
                     RRLQD"
BASE COUNT      354 a    437 c    437 g    360 t
ORIGIN
```

Query Match 84.5%; Score 1144.4; DB 40; Length 1588;
Best Local Similarity 98.2%; Pred. No. 7.9e-252;
Matches 1231; Conservative 0; Mismatches 16; Indels 7; Gaps 7;

```
Qy    100 acaaccccggaggcgcagaccgtccacacgggacagggacgcccgctcggctccgctccg 159
          | ||||| ||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db    224 AAAACCCTGGAGGCGCAGACCGTCCACACGAGACAGGGACGCCCGCTCGGCTCCGCTCCG 283

Qy    160 cgctgatcctcaaggtctcgcggtttccccgccgcccacccggacgccgacgaaagccag 219
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    284 -GCTGATCCTCAAGGTCTCGCGGTTTCCCCGCCGCCCACCCGGACGCCGACGAAAGCCAG 342

Qy    220 cgagctcctcagcctcaggcatctgcatctgggaccgacctcctgggctggctgatcaaa 279
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    343 CGAGCTCCTCAGTCTCAGGCATCTGCATCTGGGACCGACCTCCTGGGCTGGCTGATCAAA 402

Qy    280 gaggaagcagcagcaatgtctgctgtgggggctgcaactccatacctgcatcatcctggt 339
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db    403 GAGGAAGCAGCAGCAATGTCTGCTGTGGGGACTGCAACTCCATACCTGCATCATCCTGGT 462

Qy    340 gatagtcacagtggccgagtgagtttcttgggggcccagcttcctccagaggtggcagca 399
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    463 GATAGTCACAGTGGCCGAGTGAGTTTCTTGGGGGCCCAGCTTCCTCCAGAGGTGGCAGCA 522

Qy    400 atggcccggctactaggggacctagacaggagcacgttcagaaagttgctgaagtttgtg 459
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    523 ATGGCCCGGCTACTAGGGGACCTAGACAGGAGCACGTTCAGAAAGTTGCTGAAGTTTGTG 582

Qy    460 gtcagcagcctgcaggggggaggactgccgagacggtgtgcagcgtcttggggtcagcgcc 519
          |||||||||||||||||||||||||||||||| || ||||||||||||||||||||||||
```

```
Db    583 GTCAGCAGCCTGCAGGGGGAGGACTGCCGAGACTGT-TGCAGCGTCTTGGGGTCAGCGCC 641

Qy    520 aacctgccggaggagcagctgggtgccctgctggcaggcatgcacacactgctccagcag 579
          |||||||||| |||||||||||||| ||||||||||||||||||||||||||||||||||
Db    642 AACCTGCCGGA-GAGCAGCTGGGTG-CCTGCTGGCAGGCATGCACACACTGCTCCAGCAG 699

Qy    580 gccctccgtctgccccccaccagcctgaagcctgacaccttcaggga-ccagctccagga 638
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    700 GCCCTCCGTCTGCCCCCCACCAGCCTGAAGCCTGACACCTTCAGGGACCCAGCTCCAGGA 759

Qy    639 gctctgcatccccccaagacctggtcggggacttggccagcgtggtatttgggagccagcg 698
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    760 GCTCTGCATCCCCCCAAGACCTGGTCGGGGACTTGGCCAGCGTGGTATTTGGGAGCCAGCG 819

Qy    699 gcccctccttgattctgtggcccagcagcagggggcctggctgccgcatgttgctgactt 758
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    820 G-CCCTCCTTGATTCTGTGGCCCAGCAGCAGGGGGCCTGGCTGCCGCATGTTGCTGACTT 878

Qy    759 tcggtggcgggtggatgtagcaatctccaccagtgccctggctcgctccctgcagccgag 818
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    879 TCGGTGGCGGGTGGATGTAGCAATCTCCACCAGTGCCCTGGCTCGCTCCCTGCAGCCGAG 938

Qy    819 cgtcctgatgcagctgaagctttcagatgggtcagcataccgctttgaggtccccacagc 878
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    939 CGTCCTGATGCAGCTGAAGCTTTCAGATGGGTCAGCATACCGCTTTGAGGTCCCCACAGC 998

Qy    879 caagttccaggagctgcggtacagcgtggccctggtcctaaaggagatggcagatctgga 938
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    999 CAAGTTCCAGGAGCTGCGGTACAGCGTGGCCCTGGTCCTAAAGGAGATGGCAGATCTGGA 1058

Qy    939 gaagaggtgtgagcgcagactgcaggactgacccctcacttgaccagtcccattcagatc 998
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1059 GAAGAGGTGTGAGCGCAGACTGCAGGACTGACCCCTCACTTGACCAGTCCCATTCAGATC 1118

Qy    999 cggcttggacaggcacctgagatggtgccaaagtgcagctgactcttcccacgacagccc 1058
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1119 CGGCTTGGACAGGCACCTGAGATGGTGCCAAAGTGCAGCTGACTCTTCCCACGACAGCCC 1178

Qy    1059 tggccttcccatcaggcaggctcttcagtgagtgtttgaacgtaattatgtagttttctg 1118
           || ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   1179 TGCCCTTCCCATGAGGCAGGTCTTCAGTGAGTGTTTGAACGTAATTATGTAGTTTTCTG 1238

Qy    1119 tttaattgaaaaagagagctatgccttttttttcttttttggaagtaaagcagctaaaaca 1177
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1239 TTTAATTGAAAAAGAGAGCTATGCCTTTTTTTCTTTTTGGAAGTAAAGCAGCTAAAAACA 1298

Qy    1178 tgtttctataggtgagtgttggaccttcacacctccccttccctgtacatttgtctttgg 1237
           |||||||||||||||||||||||||||||     ||||||||||||||||||||||||||
Db   1299 TGTTTCTATAGGTGAGTGTTGGACCTCACACCCTCCCCTTCCCTGTACATTTGTCTTTGG 1358

Qy    1238 tgctggacgtggccatgtgaggccaggttgaggccctttgtagacaacatacagttgctc 1297
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1359 TGCTGGACGTGGCCATGTGAGGCCAGGTTGAGGCCCTTTGTAGACAACATACAGTTGCTC 1418

Qy    1298 agcctggccccatgtagccaggtgctttttgtagatcttgtgtttcaggcagggc 1351
           |||||||||||||||||||||||||||||||||||||||||||||| |||  ||||
Db   1419 AGCCTGGCCCCATGTAGCCAGGTGCTTTTGTAGATCTTGTGTTTTAGGTTGGGC 1472
```

```
RESULT   2
AF124523/c
LOCUS       AF124523   108369 bp    DNA             PRI       24-AUG-1999
DEFINITION  Homo sapiens chromosome 8 clone Kox18-21 containing zinc-finger
            protein 7 (ZFP7) gene, complete sequence.
ACCESSION   AF124523
VERSION     AF124523.1  GI:4325309
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 108369)
  AUTHORS   Schudy,A., Rump,A., Drescher,B., Koczan,D., Thiesen,H.-J.,
```

```
            Merck-Rousseau,M.F. and Rosenthal,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-1999) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES             Location/Qualifiers
     source          1. .108369
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /clone="Kox18-21"
     repeat_region   1. .245
                     /rpt_family="AluY"
                     /evidence=not_experimental
     repeat_region   246. .584
                     /rpt_family="L1MA4"
                     /evidence=not_experimental
     repeat_region   585. .863
                     /rpt_family="AluJo"
                     /evidence=not_experimental
     repeat_region   864. .1068
                     /rpt_family="L1MA4"
                     /evidence=not_experimental
     repeat_region   1069. .1353
                     /rpt_family="AluSx"
                     /evidence=not_experimental
     repeat_region   1387. .1734
                     /rpt_family="L1MA4"
                     /evidence=not_experimental
     repeat_region   1735. .2042
                     /rpt_family="AluY"
                     /evidence=not_experimental
     repeat_region   2043. .2172
                     /rpt_family="L1MA4"
                     /evidence=not_experimental
     repeat_region   2173. .2482
                     /rpt_family="AluSx"
                     /evidence=not_experimental
     repeat_region   2483. .2901
                     /rpt_family="L1MA4"
                     /evidence=not_experimental
     exon            2835. .2912
                     /note="GRAIL, score = 45.000%, comment = marginal"
                     /evidence=not_experimental
     exon            2848. .2883
                     /note="Xpound exon prediction, score = 77% (0%)"
                     /evidence=not_experimental
     repeat_region   2902. .3202
                     /rpt_family="AluSp"
                     /evidence=not_experimental
     repeat_region   3203. .3325
                     /rpt_family="L1MA4"
                     /evidence=not_experimental
     repeat_region   3326. .3625
                     /rpt_family="AluSx"
                     /evidence=not_experimental
     repeat_region   3626. .3968
                     /rpt_family="L1MA4"
                     /evidence=not_experimental
     repeat_region   4104. .4431
                     /rpt_family="L2"
                     /evidence=not_experimental
     exon            4123. .4369
                     /note="GRAIL, score = 63.000%, comment = good"
                     /evidence=not_experimental
     repeat_region   4450. .4608
                     /rpt_family="AluSg/x"
                     /evidence=not_experimental
     repeat_region   4623. .4796
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   4898. .4997
                     /rpt_family="AluSg"
                     /evidence=not_experimental
     repeat_region   5000. .5060
                     /rpt_family="L1MA4A"
                     /evidence=not_experimental
     repeat_region   5101. .5402
                     /rpt_family="AluY"
                     /evidence=not_experimental
     repeat_region   complement(5578. .5595)
                     /rpt_family="MER33"
                     /evidence=not_experimental
     repeat_region   5596. .5867
                     /rpt_family="AluSx"
                     /evidence=not_experimental
     repeat_region   5902. .5912
                     /rpt_family="AluSx"
                     /evidence=not_experimental
     repeat_region   complement(5913. .6139)
                     /rpt_family="MER33"
                     /evidence=not_experimental
     repeat_region   complement(8764. .9075)
                     /rpt_family="AluSx"
                     /evidence=not_experimental
     repeat_region   10935. .11017
                     /rpt_family="MER74B"
                     /evidence=not_experimental
     repeat_region   complement(11018. .11306)
                     /rpt_family="AluJb"
                     /evidence=not_experimental
     repeat_region   11307. .11724
                     /rpt_family="MER74B"
                     /evidence=not_experimental
     repeat_region   11725. .12006
                     /rpt_family="AluSq"
                     /evidence=not_experimental
     repeat_region   12049. .12100
                     /note="homology = 100.00%, score = 26, counts = 2"
                     /rpt_type=tandem
                     /rpt_unit=actcaccttgagatgtgggtgtctgc
                     /evidence=not_experimental
     exon            complement(12081. .12230)
                     /note="GRAIL, score = 56.000%, comment = good"
                     /evidence=not_experimental
     repeat_region   complement(12255. .12748)
                     /rpt_family="LTR42"
                     /evidence=not_experimental
     exon            14255. .14377
                     /note="GenScan, score = 10.27%, comment = Initial_exon 123
                     bp frame: 1 phase: 0"
                     /evidence=not_experimental
     exon            14386. .14580
                     /gene="zinc-finger protein 7 (ZFP7)"
                     /number=1
     mRNA            join(14386. .14580,15811. .15858,16246. .16372,23346. .23462,
                     27285. .29150)
                     /gene="zinc-finger protein 7 (ZFP7)"
                     /note="mRNA acc.no. M29580"
                     /product="ZFP7"
     gene            14386. .29150
                     /gene="zinc-finger protein 7 (ZFP7)"
     repeat_region   15176. .15595
                     /rpt_family="L2"
                     /evidence=not_experimental
     exon            15811. .15858
                     /gene="zinc-finger protein 7 (ZFP7)"
                     /number=2
     CDS             join(15856. .15858,16246. .16372,23346. .23462,27285. .29098)
                     /gene="zinc-finger protein 7 (ZFP7)"
                     /note="mRNA acc.no. M29580"
                     /codon_start=1
                     /product="zinc-finger protein 7"
                     /protein_id="AAD17326.1"
                     /db_xref="GI:4325310"
                     /translation="MEVVTFGDVAVHFSREEWQCLDPGQRALYREVMLENHSSVAGLA
                     GFLVFKPELISRLEQGEEPWVLDLQGAEGTEAPRTSKTDSTIRTENEQACEDMDILKS
```

```
                ESYGTVVRISPQDFPQNPGFGDVSDSEVWLDSHLGSPGLKVTGFTFQNNCLNEETVVP
                KTFTKDAPQGCKELGSSGLDCQPLESQGESAEGMSQRCEECGKGIRATSDIALHWEIN
                TQKISRCQECQKKLSDCLQGKHTNNCHGEKPYECAECGKVFRLCSQLNQHQRIHTGEK
                PFKCTECGKAFRLSSKLIQHQRIHTGEKPYRCEECGKAFGQSSSLIHHQRIHTGERPY
                GCRECGKAFSQQSQLVRHQRTHTGERPYPCKECGKAFSQSSTLAQHQRMHTGEKAQIL
                KASDSPSLVAHQRIHAVEKPFKCDECGKAFRWISRLSQHQLIHTGEKPYKCNKCTKAF
                GCSSRLIRHQRTHTGEKPFKCDECGKGFVQGSHLIQHQRIHTGEKPYVCNDCGKAFSQ
                SSSLIYHQRIHKGEKPYECLQCGKAFSMSTQLTIHQRVHTGERPYKCNECGKAFSQNS
                TLFQHQIIHAGVKPYECSECGKAFSRSSYLIEHQRIHTRAQWFYEYGNALEGSTFVSR
                KKVNTIKKLHQCEDCEKIFRWRSHLIIHQRIHTGEKPYKCNDCGKAFNRSSRLTQHQK
                IHMG"
    exon        16246. .16372
                /gene="zinc-finger protein 7 (ZFP7)"
                /number=3
    repeat_region   complement(17102. .17300)
                /rpt_family="L1ME"
                /evidence=not_experimental
    repeat_region   complement(17497. .17665)
                /rpt_family="AluSg/x"
                /evidence=not_experimental
    repeat_region   18270. .18401
                /rpt_family="AluYb8"
                /evidence=not_experimental
    repeat_region   18404. .18692
                /rpt_family="AluSx"
                /evidence=not_experimental
    repeat_region   18871. .19172
                /rpt_family="AluSx"
                /evidence=not_experimental
    repeat_region   complement(19300. .19632)
                /rpt_family="AluSx"
                /evidence=not_experimental
    repeat_region   19635. .19688
                /note="homology = 74.10%, score = 20, counts = 9"
                /rpt_type=tandem
                /rpt_unit=tttgtt
                /evidence=not_experimental
    repeat_region   complement(19740. .20031)
                /rpt_family="AluSp"
                /evidence=not_experimental
    repeat_region   complement(20033. .20103)
                /rpt_family="L1M4"

  Query Match             80.1%;  Score 1085.8;  DB 40;  Length 108369;
  Best Local Similarity   98.8%;  Pred. No. 2e-238;
  Matches 1104;  Conservative    0;  Mismatches   12;  Indels    1;  Gaps    1;

Qy    236 aggcatctgcatctgggaccgacctcctgggctggctgatcaaagaggaagcagcagcaa 295
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  37325 AGGCATCTGCATCTGGGACCGACCTCCTGGGCTGGCTGATCAAAGAGGAAGCAGCAGCAA 37266

Qy    296 tgtctgctgtgggggctgcaactccatacctgcatcatcctggtgatagtcacagtggcc 355
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  37265 TGTCTGCTGTGGGGGCTGCAACTCCATACCTGCATCATCCTGGTGATAGTCACAGTGGCC 37206

Qy    356 gagtgagtttcttgggggcccagcttcctccagaggtggcagcaatggcccggctactag 415
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  37205 GAGTGAGTTTCTTGGGGGCCCAGCTTCCTCCAGAGGTGGCAGCAATGGCCCGGCTACTAG 37146

Qy    416 gggacctagacaggagcacgttcagaaagttgctgaagtttgtggtcagcagcctgcagg 475
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  37145 GGGACCTAGACAGGAGCACGTTCAGAAAGTTGCTGAAGTTTGTGGTCAGCAGCCTGCAGG 37086

Qy    476 gggaggactgccgagacggtgtgcagcgtcttggggtcagcgccaacctgccggaggagc 535
          |||||||||||||||| | ||||||||||||||||||||||||||||||||||||||||
Db  37085 GGGAGGACTGCCGAGAGGCTGTGCAGCGTCTTGGGGTCAGCGCCAACCTGCCGGAGGAGC 37026

Qy    536 agctgggtgccctgctggcaggcatgcacacactgctccagcaggccctccgtctgcccc 595
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  37025 AGCTGGGTGCCCTGCTGGCAGGCATGCACACACTGCTCCAGCAGGCCCTCCGTCTGCCCC 36966

Qy    596 ccaccagcctgaagcctgacaccttcagggaccagctccaggagctctgcatccccaag 655
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36965 CCACCAGCCTGAAGCCTGACACCTTCAGGGACCAGCTCCAGGAGCTCTGCATCCCCCAAG 36906

Qy    656 acctggtcggggacttggccagcgtggtatttgggagccagcggcccctccttgattctg 715
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36905 ACCTGGTCGGGGACTTGGCCAGCGTGGTATTTGGGAGCCAGCGGCCCCTCCTTGATTCTG 36846

Qy    716 tggcccagcagcaggggggcctggctgccgcatgttgctgactttcggtggcgggtggatg 775
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36845 TGGCCCAGCAGCAGGGGGGCCTGGCTGCCGCATGTTGCTGACTTTCGGTGGCGGGTGGATG 36786

Qy    776 tagcaatctccaccagtgccctggctcgctccctgcagccgagcgtcctgatgcagctga 835
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36785 TAGCAATCTCCACCAGTGCCCTGGCTCGCTCCCTGCAGCCGAGCGTCCTGATGCAGCTGA 36726

Qy    836 agctttcagatgggtcagcataccgctttgaggtccccacagccaagttccaggagctgc 895
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36725 AGCTTTCAGATGGGTCAGCATACCGCTTTGAGGTCCCCACAGCCAAGTTCCAGGAGCTGC 36666

Qy    896 ggtacagcgtggccctggtcctaaaggagatggcagatctggagaagaggtgtgagcgca 955
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36665 GGTACAGCGTGGCCCTGGTCCTAAAGGAGATGGCAGATCTGGAGAAGAGGTGTGAGCGCA 36606

Qy    956 gactgcaggactgacccctcacttgaccagtcccattcagatccggcttggacaggcacc 1015
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36605 GACTGCAGGACTGACCCCTCACTTGACCAGTCCCATTCAGATCCGGCTTGGACAGGCACC 36546

Qy   1016 tgagatggtgccaaagtgcagctgactcttcccacgacagccctggccttcccatcaggc 1075
          |||||||||||||||||||||||||||||||||||||||||||||| |||||||| ||||
Db  36545 TGAGATGGTGCCAAAGTGCAGCTGACTCTTCCCACGACAGCCCTGCCCTTCCCATGAGGC 36486

Qy   1076 aggctcttcagtgagtgtttgaacgtaattatgtagttttctgtttaattgaaaaagaga 1135
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36485 AGGCTCTTCAGTGAGTGTTTGAACGTAATTATGTAGTTTTCTGTTTAATTGAAAAAGAGA 36426

Qy   1136 gctatgccttttttttcttttttggaagtaaagcagct-aaaacatgtttctataggtgagt 1194
          |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db  36425 GCTATGCCTTTTTTTCTTTTTGGAAGTAAAGCAGCTAAAAACATGTTTCTATAGGTGAGT 36366

Qy   1195 gttggaccttcacacctccccttccctgtacatttgtctttggtgctggacgtggccatg 1254
          |||||||||     ||||||||||||||||||||||||||||||||||||||||||||||
Db  36365 GTTGGACCTCACACCCTCCCCTTCCCTGTACATTTGTCTTTGGTGCTGGACGTGGCCATG 36306

Qy   1255 tgaggccaggttgaggccctttgtagacaacatacagttgctcagcctggccccatgtag 1314
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36305 TGAGGCCAGGTTGAGGCCCTTTGTAGACAACATACAGTTGCTCAGCCTGGCCCCATGTAG 36246

Qy   1315 ccaggtgcttttgtagatcttgtgtttcaggcagggc 1351
          |||||||||||||||||||||||||| |||  ||||
Db  36245 CCAGGTGCTTTTGTAGATCTTGTGTTTTAGGTTGGGC 36209


RESULT  3
AF216665/c
LOCUS       AF216665   174923 bp    DNA             HTG       30-JAN-2000
DEFINITION  Homo sapiens chromosome 8 clone BAC 704B241056 map 8q24, ***
            SEQUENCING IN PROGRESS ***, in unordered pieces.
ACCESSION   AE216665
VERSION     AF216665.1  GI:6815808
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 174923)
  AUTHORS   Schudy,A., Schilhabel,M., Baumgart,C., Menzel,U., Weber,J.,
            Schattevoy,R. and Rosenthal,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-DEC-1999) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
COMMENT     1-75864: contig of 75864 bp;  75864-75865: gap of unknown size;
            75865-109038: contig of 33174 bp;  109038-109039: gap of unknown
            size;  109039-117368: contig of 8330 bp;  117368-117369: gap of
            unknown size;  117369-128688: contig of 11320 bp;  128688-128689:
```

```
            gap of unknown size;  128689-134145: contig of 5457 bp;
            134145-134146: gap of unknown size;  134146-147620: contig of 13475
            bp;  147620-147621: gap of unknown size;  147621-149478: contig of
            1858 bp;  149478-149479: gap of unknown size;  149479-151777:
            contig of 2299 bp;  151777-151778: gap of unknown size;
            151778-155130: contig of 3353 bp;  155130-155131: gap of unknown
            size;  155131-157254: contig of 2124 bp;  157254-157255: gap of
            unknown size;  157255-158492: contig of 1238 bp;  158492-158493:
            gap of unknown size;  158493-174923: contig of 16431 bp;.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1. .174923
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8q24"
                     /clone="BAC 704B241056"
     repeat_region   complement(36. .349)
                     /rpt_family="AluSx"
     repeat_region   352. .405
                     /rpt_type=tandem
     exon            complement(449. .582)
                     /note="FEXHB"
                     /evidence=not_experimental
     repeat_region   complement(457. .748)
                     /rpt_family="AluSq"
     repeat_region   complement(750. .820)
                     /rpt_family="L1M4_orf2"
     repeat_region   complement(1167. .1754)
                     /rpt_family="L1M4_orf2"
     repeat_region   1387. .1410
                     /rpt_type=tandem
     repeat_region   complement(1410. .1713)
                     /rpt_family="AluY"
     exon            1434. .1555
                     /note="FEXHB"
                     /evidence=not_experimental
     repeat_region   1517. .1702
                     /rpt_type=inverted
     repeat_region   complement(1865. .2170)
                     /rpt_family="AluJo"
     repeat_region   2386. .2461
                     /rpt_type=tandem
     repeat_region   complement(2435. .2544)
                     /rpt_family="AluJb"
     repeat_region   complement(2545. .2649)
                     /rpt_family="AluSx"
     repeat_region   2738. .3017
                     /rpt_family="AluJo"
     repeat_region   2747. .2931
                     /rpt_type=inverted
     exon            3369. .3458
                     /note="GENSCAN"
     exon            4066. .4182
                     /note="GENSCAN"
     exon            4066. .4182
                     /note="MZEF"
                     /evidence=not_experimental
     repeat_region   complement(4603. .5038)
                     /rpt_family="MLT1C"
     exon            complement(4836. .4889)
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            complement(5199. .5235)
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            complement(5199. .5235)
                     /note="MZEF"
                     /evidence=not_experimental
     repeat_region   5336. .5612
                     /rpt_family="AluY"
     repeat_region   5472. .5530
                     /rpt_type=inverted
     repeat_region   5627. .5906
                     /rpt_type=tandem
     repeat_region   complement(6589. .6751)
                     /rpt_family="AluSg"
     repeat_region   6689. .6747
                     /rpt_type=inverted
     exon            8005. .8268
                     /note="MZEF"
                     /evidence=not_experimental
     exon            8005. .9818
                     /note="GENSCAN"
     exon            8046. .9818
                     /note="FEXHB"
                     /evidence=not_experimental
     exon            8046. .9818
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            complement(9207. .9412)
                     /note="GRAIL"
                     /evidence=not_experimental
     repeat_region   9299. .9578
                     /rpt_type=tandem
     repeat_region   9958. .10259
                     /rpt_family="AluY"
     exon            complement(10113. .10232)
                     /note="FEXHB"
                     /evidence=not_experimental
     repeat_region   10238. .10259
                     /rpt_type=tandem
     repeat_region   11078. .11384
                     /rpt_family="AluY"
     repeat_region   11359. .11384
                     /rpt_type=tandem
     repeat_region   11519. .11638
                     /rpt_type=tandem
     repeat_region   complement(11792. .11848)
                     /rpt_family="LTR40b"
     exon            complement(11834. .11860)
                     /note="GENSCAN"
     repeat_region   complement(12208. .12332)
                     /rpt_family="HERVL40"
     exon            complement(12368. .12424)
                     /note="GENSCAN"
     exon            complement(12368. .12424)
                     /note="XPOUND"
                     /evidence=not_experimental
     exon            complement(12368. .12468)
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            complement(12673. .12773)
                     /note="GRAIL"
                     /evidence=not_experimental
     repeat_region   12872. .12961
                     /rpt_family="HERVL40"
     repeat_region   12985. .13069
                     /rpt_family="MER58B"
     repeat_region   13354. .13491
                     /rpt_type=inverted
     repeat_region   complement(13354. .13641)
                     /rpt_family="AluY"
     repeat_region   13949. .14111
                     /rpt_family="FRAM"
     repeat_region   14263. .14357
                     /rpt_family="L2b"
     exon            complement(14534. .14650)
                     /note="GRAIL"
                     /evidence=not_experimental
     repeat_region   14715. .14969
                     /rpt_family="AluJo"
     repeat_region   15002. .15162
```

```
                     /rpt_family="AluSq"
     repeat_region   15022. .15161
                     /rpt_type=inverted
     repeat_region   15168. .15312
                     /rpt_family="MER58A"
     repeat_region   16383. .16634
                     /rpt_family="L2"
     exon            16705. .16831
                     /note="GRAIL"
                     /evidence=not_experimental
     repeat_region   16714. .16761
                     /rpt_family="L2a"
     repeat_region   16722. .16761
                     /rpt_family="L2b"
     repeat_region   16782. .16912
                     /rpt_family="L2a"
     exon            complement(17315. .17989)
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            complement(17315. .18046)
                     /note="GENSCAN"
     exon            complement(17321. .17848)
                     /note="XPOUND"
                     /evidence=not_experimental
     exon            complement(17338. .17681)
                     /note="MZEF"
                     /evidence=not_experimental
     exon            17351. .17527
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            17537. .17863
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            complement(18086. .18181)

  Query Match             80.0%;  Score 1084.2;  DB 52;  Length 174923;
  Best Local Similarity   98.7%;  Pred. No. 4.7e-238;
  Matches 1103;  Conservative    0;  Mismatches   13;  Indels    1;  Gaps    1;

Qy    236 aggcatctgcatctgggaccgacctcctgggctggctgatcaaagaggaagcagcagcaa 295
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18048 AGGCATCTGCATCTGGGACCGACCTCCTGGGCTGGCTGATCAAAGAGGAAGCAGCAGCAA 17989

Qy    296 tgtctgctgtgggggctgcaactccatacctgcatcatcctggtgatagtcacagtggcc 355
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17988 TGTCTGCTGTGGGGGCTGCAACTCCATACCTGCATCATCCTGGTGATAGTCACAGTGGCC 17929

Qy    356 gagtgagtttcttgggggcccagcttcctccagaggtggcagcaatggcccggctactag 415
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17928 GAGTGAGTTTCTTGGGGGCCCAGCTTCCTCCAGAGGTGGCAGCAATGGCCCGGCTACTAG 17869

Qy    416 gggacctagacaggagcacgttcagaaagttgctgaagtttgtggtcagcagcctgcagg 475
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17868 GGGACCTAGACAGGAGCACGTTCAGAAAGTTGCTGAAGTTTGTGGTCAGCAGCCTGCAGG 17809

Qy    476 gggaggactgccgagacggtgtgcagcgtcttggggtcagcgccaacctgccggaggagc 535
          |||||||||||||||| | ||||||||||||||||||||||||||||||||||||||||||
Db  17808 GGGAGGACTGCCGAGAGGCTGTGCAGCGTCTTGGGGTCAGCGCCAACCTGCCGGAGGAGC 17749

Qy    536 agctgggtgccctgctggcaggcatgcacacactgctccagcaggccctccgtctgcccc 595
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17748 AGCTGGGTGCCCTGCTGGCAGGCATGCACACACTGCTCCAGCAGGCCCTCCGTCTGCCCC 17689

Qy    596 ccaccagcctgaagcctgacaccttcagggaccagctccaggagctctgcatccccccaag 655
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17688 CCACCAGCCTGAAGCCTGACACCTTCAGGGACCAGCTCCAGGAGCTCTGCATCCCCCAAG 17629

Qy    656 acctggtcggggacttggccagcgtggtatttgggagccagcggcccctccttgattctg 715
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17628 ACCTGGTCGGGGACTTGGCCAGCGTGGTATTTGGGAGCCAGCGGCCCCTCCTTGATTCTG 17569

Qy    716 tggcccagcagcaggggggcctggctgccgcatgttgctgactttcggtggcgggtggatg 775
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db  17568 TGGCCCAGCAGCAGGGGGCCTGGCTGCCGCATGTTGCTGACTTTCGGTGGCGGGTGGATG 17509

Qy    776 tagcaatctccaccagtgccctggctcgctccctgcagccgagcgtcctgatgcagctga 835
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17508 TAGCAATCTCCACCAGTGCCCTGGCTCGCTCCCTGCAGCCGAGCGTCCTGATGCAGCTGA 17449

Qy    836 agctttcagatgggtcagcataccgctttgaggtccccacagccaagttccaggagctgc 895
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17448 AGCTTTCAGATGGGTCAGCATACCGCTTTGAGGTCCCCACAGCCAAGTTCCAGGAGCTGC 17389

Qy    896 ggtacagcgtggccctggtcctaaaggagatggcagatctggagaagaggtgtgagcgca 955
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17388 GGTACAGCGTGGCCCTGGTCCTAAAGGAGATGGCAGATCTGGAGAAGAGGTGTGAGCGCA 17329

Qy    956 gactgcaggactgacccctcacttgaccagtcccattcagatccggcttggacaggcacc 1015
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17328 GACTGCAGGACTGACCCCTCACTTGACCAGTCCCATTCAGATCCGGCTTGGACAGGCACC 17269

Qy   1016 tgagatggtgccaaagtgcagctgactcttcccacgacagccctggccttcccatcaggc 1075
          |||||||||||||||||||||||||||||||||||||||||| |||||||| ||||
Db  17268 TGAGATGGTGCCAAAGTGCAGCTGACTCTTCCCACGACAGCCCTGCCCTTCCCATGAGGC 17209

Qy   1076 aggctcttcagtgagtgtttgaacgtaattatgtagttttctgtttaattgaaaaagaga 1135
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17208 AGGCTTTTCAGTGAGTGTTTGAACGTAATTATGTAGTTTTCTGTTTAATTGAAAAAGAGA 17149

Qy   1136 gctatgccttttttttctttttggaagtaaagcagct-aaaacatgtttctataggtgagt 1194
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db  17148 GCTATGCCTTTTTTTCTTTTTGGAAGTAAAGCAGCTAAAAACATGTTTCTATAGGTGAGT 17089

Qy   1195 gttggaccttcacacctccccttccctgtacatttgtctttggtgctggacgtggccatg 1254
          |||||||||     ||||||||||||||||||||||||||||||||||||||||||||
Db  17088 GTTGGACCTCACACCCTCCCCTTCCCTGTACATTTGTCTTTGGTGCTGGACGTGGCCATG 17029

Qy   1255 tgaggccaggttgaggccctttgtagacaacatacagttgctcagcctggccccatgtag 1314
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17028 TGAGGCCAGGTTGAGGCCCTTTGTAGACAACATACAGTTGCTCAGCCTGGCCCCATGTAG 16969

Qy   1315 ccaggtgcttttgtagatcttgtgtttcaggcagggc 1351
          |||||||||||||||||||||||||| |||  ||||
Db  16968 CCAGGTGCTTTTGTAGATCTTGTGTTTTAGGTTGGGC 16932


RESULT   4
AF118808
LOCUS       AF118808   128009 bp    DNA             PRI       30-JUN-1999
DEFINITION  Homo sapiens chromosome 8qter PAC clone P2G2, complete sequence.
ACCESSION   AF118808
VERSION     AF118808.1  GI:4337002
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoan; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 128009)
  AUTHORS   Rump,A., Rosenthal,A., Drescher,B., Weber,J., Schattevoy,R.,
            Koczan,D., Thiesen,H.-J. and Merck-Rousseau,M.-F.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JAN-1999) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES             Location/Qualifiers
     source          1. .128009
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8qter"
                     /clone="PAC clone P2G2"
                     /note="mapped to 8qter by FISH; identified by filter
                     hybridization with zinc finger specific probe: clone
                     contains two sequencing gaps (60562-60589 and
                     68689-68687); orientation of the middle contig is not
                     known"
     exon            complement(2184. .2219)
```

```
                     /note="Exon predicted by several exon prediction programs
                     and by homology to GenBank entry HUMZIFIA_1: Homo sapiens
                     zinc finger protein mRNA, 3' end."
                     /evidence=not_experimental
     exon            2269. .2310
                     /note="Exon predicted by several exon prediction programs
                     and by homology to GenBank entry HUMZNFN_1: Homo sapiens
                     DNA-binding protein (ZNF) gene, partial cds."
                     /evidence=not_experimental
     exon            47097. .47263
                     /note="Exon predicted by several exon prediction programs
                     and by homology to GenBank entry HSKOX9_1: Human Kox9 mRNA
                     for zinc finger protein, partial."
                     /evidence=not_experimental
     exon            47430. .47631
                     /note="Exon predicted by several exon prediction programs
                     and by homology to GenBank entry AF024700_1: Homo sapiens
                     clone B4-2 zinc finger protein mRNA, partial cds."
                     /evidence=not_experimental
     exon            94802. .95455
                     /note="Exon predicted by several exon prediction programs
                     and by homology to GenBank entry HSZFP647_1: Human mRNA
                     for zinc finger protein (clone 647)."
                     /evidence=not_experimental
     exon            complement(95143. .95187)
                     /note="Exon predicted by several exon prediction programs
                     and by homology to GenBank entry HSZNFPT7_1: H.sapiens
                     mRNA for Zinc-finger protein (ZNFpT7)."
                     /evidence=not_experimental
BASE COUNT    35117 a  30514 c  28979 g  33352 t     47 others
ORIGIN

  Query Match           79.8%;  Score 1081;  DB 40;  Length 128009;
  Best Local Similarity  98.6%;  Pred. No. 2.6e-237;
  Matches 1101;  Conservative    0;  Mismatches   15;  Indels    1;  Gaps    1;

Qy     236 aggcatctgcatctgggaccgacctcctgggctggctgatcaaagaggaagcagcagcaa 295
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  125842 AGGCATCTGCATCTGGGACCGACCTCCTGGGCTGGCTGATCAAAGAGGAAGCAGCAGCAA 125901

Qy     296 tgtctgctgtgggggctgcaactccatacctgcatcatcctggtgatagtcacagtggcc 355
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  125902 TGTCTGCTGTGGGGGCTGCAACTCCATACCTGCATCATCCTGGTGATAGTCACAGTGGCC 125961

Qy     356 gagtgagtttcttgggggcccagcttcctccagaggtggcagcaatggcccggctactag 415
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db  125962 GAGTGAGTTTCTTGGGGGCCCAGCTTCCTCCAAAGGTGGCAGCAATGGCCCGGCTACTAG 126021

Qy     416 gggacctagacaggagcacgttcagaaagttgctgaagtttgtggtcagcagcctgcagg 475
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  126022 GGGACCTAGACAGGAGCACGTTCAGAAAGTTGCTGAAGTTTGTGGTCAGCAGCCTGCAGG 126081

Qy     476 gggaggactgccgagacggtgtgcagcgtcttggggtcagcgccaacctgccggaggagc 535
           |||||||||||||| | | |||||| ||||||||||||||||||||||||||||||||||
Db  126082 GGGAGGACTGCCGAAAGGCTGTGCACCGTCTTGGGGTCAGCGCCAACCTGCCGGAGGAGC 126141

Qy     536 agctgggtgccctgctggcaggcatgcacacactgctccagcaggccctccgtctgcccc 595
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  126142 AGCTGGGTGCCCTGCTGGCAGGCATGCACACACTGCTCCAGCAGGCCCTCCGTCTGCCCC 126201

Qy     596 ccaccagcctgaagcctgacaccttcagggaccagctccaggagctctgcatccccccaag 655
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  126202 CCACCAGCCTGAAGCCTGACACCTTCAGGGACCAGCTCCAGGAGCTCTGCATCCCCCAAG 126261

Qy     656 acctggtcggggacttggccagcgtggtatttgggagccagcggcccctccttgattctg 715
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  126262 ACCTGGTCGGGGACTTGGCCAGCGTGGTATTTGGGAGCCAGCGGCCCCTCCTTGATTCTG 126321

Qy     716 tggcccagcagcaggggcctggctgccgcatgttgctgactttcggtggcgggtggatg 775
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  126322 TGGCCCAGCAGCAGGGGGCCTGGCTGCCGCATGTTGCTGACTTTCGGTGGCGGGTGGATG 126381

Qy     776 tagcaatctccaccagtgccctggctcgctccctgcagccgagcgtcctgatgcagctga 835
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  126382 TAGCAATCTCCACCAGTGCCCTGGCTCGCTCCCTGCAGCCGAGCGTCCTGATGCAGCTGA 126441

Qy     836 agctttcagatgggtcagcataccgctttgaggtccccacagccaagttccaggagctgc 895
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  126442 AGCTTTCAGATGGGTCAGCATACCGCTTTGAGGTCCCCACAGCCAAGTTCCAGGAGCTGC 126501

Qy     896 ggtacagcgtggccctggtcctaaaggagatggcagatctggagaagaggtgtgagcgca 955
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  126502 GGTACAGCGTGGCCCTGGTCCTAAAGGAGATGGCAGATCTGGAGAAGAGGTGTGAGCGCA 126561

Qy     956 gactgcaggactgacccctcacttgaccagtcccattcagatccggcttggacaggcacc 1015
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  126562 GACTGCAGGACTGACCCCTCACTTGACCAGTCCCATTCAGATCCGGCTTGGACAGGCACC 126621

Qy    1016 tgagatggtgccaaagtgcagctgactcttcccacgacagccctggccttcccatcaggc 1075
           |||||||||||||||||||||||||||||||||||||||| ||||||||| ||||
Db  126622 TGAGATGGTGCCAAAGTGCAGCTGACTCTTCCCACGACAGCCCTGCCCTTCCCATGAGGC 126681

Qy    1076 aggctcttcagtgagtgtttgaacgtaattatgtagttttctgtttaattgaaaaagaga 1135
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  126682 AGGCTCTTCAGTGAGTGTTTGAACGTAATTATGTAGTTTTCTGTTTAATTGAAAAAGAGA 126741

Qy    1136 gctatgccttttttttctttt ggaagtaaagcagct-aaaacatgtttctataggtgagt 1194
           |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db  126742 GCTATGCCTTTTTTTCTTTTTGGAAGTAAAGCAGCTAAAAACATGTTTCTATAGGTGAGT 126801

Qy    1195 gttggaccttcacacctccccttccctgtacatttgtctttggtgctggacgtggccatg 1254
           ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  126802 GTTGGACCTCACACCCTCCCCTTCCCTGTACATTTGTCTTTGGTGCTGGACGTGGCCATG 126861

Qy    1255 tgaggccaggttgaggccctttgtagacaacatacagttgctcagcctggccccatgtag 1314
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  126862 TGAGGCCAGGTTGAGGCCCTTTGTAGACAACATACAGTTGCTCAGCCTGGCCCCATGTAG 126921

Qy    1315 ccaggtgcttttgtagatcttgtgtttcaggcagggc 1351
           |||||||||||||||||||||||||| ||| ||||
Db  126922 CCAGGTGCTTTTGTAGATCTTGTGTTTTAGGTTGGGC 126958


RESULT   5
AF146367
LOCUS       AF146367    69528 bp    DNA             PRI       20-MAY-1999
DEFINITION  Homo sapiens chromosome 8 PAC P5A4.00 containing gene for zinc
            finger protein, complete sequence.
ACCESSION   AF146367
VERSION     AF146367.1  GI:4877991
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 69528)
  AUTHORS   Rump,A., Schudy,A., Drescher,B., Koczan,D., Thiessen,H.-J.,
            Merck-Rousseau,M.-F. and Rosenthal,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-APR-1999) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES             Location/Qualifiers
     source          1. .69528
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /clone="PAC P5A4.00"
     gene            5504. .6968
                     /gene="hypothetical"
                     /note="GenScan, score = 1186%, comment = P1_377_aa 1. .208
                     of owl|P15622|ZF64_HUMAN ZINC FINGER PROTEIN CLONE 647
                     (FRAGMENT). - HOMO SAPIENS P = 1.2e-120 S = 1186"
                     /evidence=not_experimental
     exon            5504. .6637
                     /gene="hypothetical"
```

```
                /note="GenScan, score = 117.75%, comment =
                Single-exon_gene 1134 bp frame: 1 phase: 0"
                /evidence=not_experimental
gene            7855. .44516
                /gene="hypothetical"
                /note="GenScan, score = 0%, comment = P2_471_aa GENSCAN
                predicted peptide is unknown ->NN (DEF2) prediction for
                sec struc composition: all beta sheet; overall class:
                lectin"
                /evidence=not_experimental
exon            9896. .10003
                /gene="hypothetical"
                /note="GenScan, score = -0.43%, comment = Initial_exon 108
                bp frame: 1 phase: 0"
                /evidence=not_experimental
exon            20029. .20093
                /gene="hypothetical"
                /note="GenScan, score = 0.51%, comment = Internal_exon 65
                bp frame: 0 phase: 2"
                /evidence=not_experimental
exon            20121. .20324
                /gene="hypothetical"
                /note="GenScan, score = 0.84%, comment = Internal_exon 204
                bp frame: 0 phase: 0"
                /evidence=not_experimental
exon            26797. .26852
                /gene="hypothetical"
                /note="GenScan, score = 0.81%, comment = Internal_exon 56
                bp frame: 1 phase: 2"
                /evidence=not_experimental
exon            27802. .27935
                /gene="hypothetical"
                /note="GenScan, score = 4.67%, comment = Internal_exon 134
                bp frame: 2 phase: 2"
                /evidence=not_experimental
exon            31554. .31670
                /gene="hypothetical"
                /note="GenScan, score = 0.77%, comment = Internal_exon 117
                bp frame: 2 phase: 0"
                /evidence=not_experimental
exon            36757. .37488
                /gene="hypothetical"
                /note="GenScan, score = 71.21%, comment = Terminal_exon
                732 bp frame: 0 phase: 0"
                /evidence=not_experimental
gene            complement(44951. .60537)
                /gene="KOX4"
                /note="GenScan, score = 3534%, comment = P3_742_aa 44. .686
                of owl|P17097|ZN07_HUMAN ZINC FINGER PROTEIN 7 (ZINC
                FINGER PROTEIN KOX4) (ZINC P = 0.0 S = 3534"
                /evidence=not_experimental
exon            complement(44982. .46795)
                /gene="KOX4"
                /note="GenScan, score = 113.55%, comment = Terminal_exon
                1814 bp frame: 2 phase: 2"
                /evidence=not_experimental
exon            complement(50641. .50757)
                /gene="KOX4"
                /note="GenScan, score = 7.44%, comment = Internal_exon 117
                bp frame: 1 phase: 0"
                /evidence=not_experimental
exon            complement(51365. .51454)
                /gene="KOX4"
                /note="GenScan, score = 1.36%, comment = Internal_exon 90
                bp frame: 2 phase: 0"
                /evidence=not_experimental
exon            complement(56929. .57014)
                /gene="KOX4"
                /note="GenScan, score = 6.06%, comment = Internal_exon 86
                bp frame: 1 phase: 2"
                /evidence=not_experimental
exon            complement(60537. .60658)
                /note="GenScan, score = 10.62%, comment = Initial_exon 122
                bp frame: 1 phase: 2"
                /evidence=not_experimental
BASE COUNT    17500 a  18304 c  16802 g  16922 t
ORIGIN

  Query Match             79.5%;  Score 1077.8;  DB 40;  Length 69528;
  Best Local Similarity   98.4%;  Pred. No. 1.4e-236;
  Matches 1099;  Conservative    0;  Mismatches   17;  Indels    1;  Gaps    1;

Qy    236 aggcatctgcatctgggaccgacctcctgggctggctgatcaaagaggaagcagcagcaa 295
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36755 AGGCATCTGCATCTGGGACCGACCTCCTGGGCTGGCTGATCAAAGAGGAAGCAGCAGCAA 36814

Qy    296 tgtctgctgtgggggctgcaactccatacctgcatcatcctggtgatagtcacagtggcc 355
          ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db  36815 TGTCTGCTGTGGGGACTGCAACTCCATACCTGCATCATCCTGGTGATAGTCACAGTGGCC 36874

Qy    356 gagtgagtttcttgggggcccagcttcctccagaggtggcagcaatggcccggctactag 415
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36875 GAGTGAGTTTCTTGGGGGCCCAGCTTCCTCCAGAGGTGGCAGCAATGGCCCGGCTACTAG 36934

Qy    416 gggacctagacaggagcacgttcagaaagttgctgaagtttgtggtcagcagcctgcagg 475
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36935 GGGACCTAGACAGGAGCACGTTCAGAAAGTTGCTGAAGTTTGTGGTCAGCAGCCTGCAGG 36994

Qy    476 gggaggactgccgagacggtgtgcagcgtcttggggtcagcgccaacctgccggaggagc 535
          |||||||||||||||| | ||||||||||||||||||||||||||||||||||||||||
Db  36995 GGGAGGACTGCCGAGAGGCTGTGCAGCGTCTTGGGGTCAGCGCCAACCTGCCGGAGGAGC 37054

Qy    536 agctgggtgccctgctggcaggcatgcacacactgctccagcaggccctccgtctgcccc 595
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  37055 AGCTGGGTGCCCTGCTGGCAGGCATGCACACACTGCTCCAGCAGGCCCTCCGTCTGCCCC 37114

Qy    596 ccaccagcctgaagcctgacaccttcagggaccagctccaggagctctgcatcccccaag 655
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  37115 CCACCAGCCTGAAGCCTGACACCTTCAGGGACCAGCTCCAGGAGCTCTGCATCCCCCAAG 37174

Qy    656 acctggtcggggacttggccagcgtggtatttgggagccagcggcccctccttgattctg 715
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  37175 ACCTGGTCGGGGACTTGGCCAGCGTGGTATTTGGGAGCCAGCGGCCCCTCCTTGATTCTG 37234

Qy    716 tggcccagcagcaggggggcctggctgccgcatgttgctgactttcggtggcgggtggatg 775
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  37235 TGGCCCAGCAGCAGGGGGCCTGGCTGCCGCATGTTGCTGACTTTCGGTGGCGGGTGGATG 37294

Qy    776 tagcaatctccaccagtgccctggctcgctccctgcagccgagcgtcctgatgcagctga 835
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  37295 TAGCAATTTCCACCAGTGCCCTGGCTCGCTCCCTGCAGCCGAGCGTCCTGATGCAGCTGA 37354

Qy    836 agctttcagatgggtcagcataccgctttgaggtccccacagccaagttccaggagctgc 895
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  37355 AGCTTTCAGATGGGTCAGCATACCGCTTTGAGGTCCCCACAGCCAAGTTCCAGGAGCTGC 37414

Qy    896 ggtacagcgtggccctggtcctaaaggagatggcagatctggagaagaggtgtgagcgca 955
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  37415 GGTACAGCGTGGCCCTGGTCCTAAAGGAGATGGCAGATCTGGAGAAGAGGTGTGAGCGCA 37474

Qy    956 gactgcaggactgacccctcacttgaccagtcccattcagatccggcttggacaggcacc 1015
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  37475 GACTGCAGGACTGACCCCTCACTTGACCAGTCCCATTCAGATCCGGCTTGGACAGGCACC 37534

Qy   1016 tgagatggtgccaaagtgcagctgactcttcccacgacagccctggccttcccatcaggc 1075
          |||||||||||||||||||||||||||||||||||||||||| |||||||| ||||
Db  37535 TGAGATGGTGCCAAAGTGCAGCTGACTCTTCCCACGACAGCCCTGCCCTTCCCATGAGGC 37594

Qy   1076 aggctcttcagtgagtgtttgaacgtaattatgtagttttctgtttaattgaaaaagaga 1135
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  37595 AGGCTCTTCAGTGAGTGTTTGAACGTAATTATGTAGTTTTCTGTTTAATTGAAAAAGAGA 37654

Qy   1136 gctatgccttttttttcttttttggaagtaaagcagct-aaaacatgtttctataggtgagt 1194
          |||||||||||||||||||||||||||||||||||| |||||||||||| |||||||||
Db  37655 GCTATGCCTTTTTTTTCTTTTTGGAAGTAAAGCAGCTAAAAACATGTTTCTCTAGGTGAGT 37714
```

```
Qy   1195 gttggaccttcacacctccccttccctgtacatttgtctttggtgctggacgtggccatg 1254
          ||||||||     ||||||||||||||||||||| ||||||||||||||||||||
Db  37715 GTTGGACCTCACACCCTCCCCTTCCCTGTACATTTGTTTTTGGTGCTGGACGTGGCCATG 37774

Qy   1255 tgaggccaggttgaggccctttgtagacaacatacagttgctcagcctggccccatgtag 1314
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  37775 TGAGGCCAGGTTGAGGCCCTTTGTAGACAACATACAGTTGCTCAGCCTGGCCCCATGTAG 37834

Qy   1315 ccaggtgctttttgtagatcttgtgtttcaggcagggc 1351
          ||||||||||||||| |||||||||||  |||  ||||
Db  37835 CCAGGTGCTTTTGTAAATCTTGTGTTTTAGGTTGGGC 37871


RESULT   6
AC011951/c
LOCUS       AC011951   133200 bp    DNA             HTG       13-DEC-1999
DEFINITION  Homo sapiens clone RP11-18M17, *** SEQUENCING IN PROGRESS ***, 53
            unordered pieces.
ACCESSION   AC011951
VERSION     AC011951.2  GI:6554541
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 133200)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-18M17
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 133200)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Dec 10, 1999 this sequence version replaced gi:6056232.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L3701
                Center clone name: 18_M_17
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 53 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1041: contig of 1041 bp in length
            *                   gap of unknown length
            *     1042     1130: contig of 89 bp in length
            *                   gap of unknown length
            *     1131     2412: contig of 1282 bp in length
            *                   gap of unknown length
            *     2413     3457: contig of 1045 bp in length
            *                   gap of unknown length
            *     3458     4573: contig of 1116 bp in length
            *                   gap of unknown length
            *     4574     5792: contig of 1219 bp in length
            *                   gap of unknown length
            *     5793     6853: contig of 1061 bp in length
            *                   gap of unknown length
            *     6854     8428: contig of 1575 bp in length
            *                   gap of unknown length
            *     8429     9591: contig of 1163 bp in length
            *                   gap of unknown length
            *     9592    10969: contig of 1378 bp in length
            *                   gap of unknown length
            *    10970    12475: contig of 1506 bp in length
            *                   gap of unknown length
            *    12476    14088: contig of 1613 bp in length
            *                   gap of unknown length
            *    14089    15243: contig of 1155 bp in length
            *                   gap of unknown length
            *    15244    16417: contig of 1174 bp in length
            *                   gap of unknown length
            *    16418    17553: contig of 1136 bp in length
            *                   gap of unknown length
            *    17554    18961: contig of 1408 bp in length
            *                   gap of unknown length
            *    18962    20529: contig of 1568 bp in length
            *                   gap of unknown length
            *    20530    23589: contig of 3060 bp in length
            *                   gap of unknown length
            *    23590    25249: contig of 1660 bp in length
            *                   gap of unknown length
            *    25250    26988: contig of 1739 bp in length
            *                   gap of unknown length
            *    26989    28426: contig of 1438 bp in length
            *                   gap of unknown length
            *    28427    29981: contig of 1555 bp in length
            *                   gap of unknown length
            *    29982    31552: contig of 1571 bp in length
            *                   gap of unknown length
            *    31553    33219: contig of 1667 bp in length
            *                   gap of unknown length
            *    33220    35694: contig of 2475 bp in length
            *                   gap of unknown length
            *    35695    37968: contig of 2274 bp in length
            *                   gap of unknown length
            *    37969    39872: contig of 1904 bp in length
            *                   gap of unknown length
            *    39873    42198: contig of 2326 bp in length
            *                   gap of unknown length
            *    42199    43732: contig of 1534 bp in length
            *                   gap of unknown length
            *    43733    45927: contig of 2195 bp in length
            *                   gap of unknown length
            *    45928    47914: contig of 1987 bp in length
            *                   gap of unknown length
            *    47915    50000: contig of 2086 bp in length
            *                   gap of unknown length
            *    50001    51834: contig of 1834 bp in length
            *                   gap of unknown length
            *    51835    54069: contig of 2235 bp in length
            *                   gap of unknown length
            *    54070    56363: contig of 2294 bp in length
            *                   gap of unknown length
            *    56364    59355: contig of 2991 bp in length
            *                   gap of unknown length
            *    59355    63071: contig of 3717 bp in length
            *                   gap of unknown length
            *    63072    65969: contig of 2898 bp in length
            *                   gap of unknown length
            *    65970    69646: contig of 3677 bp in length
```

```
            *                    gap of unknown length
            *    69647    73243: contig of 3597 bp in length
            *                    gap of unknown length
            *    73244    76844: contig of 3601 bp in length
            *                    gap of unknown length
            *    76845    79797: contig of 2953 bp in length
            *                    gap of unknown length
            *    79798    83127: contig of 3330 bp in length
            *                    gap of unknown length
            *    83128    85982: contig of 2855 bp in length
            *                    gap of unknown length
            *    85983    89291: contig of 3309 bp in length
            *                    gap of unknown length
            *    89292    92357: contig of 3066 bp in length
            *                    gap of unknown length
            *    92358    95772: contig of 3415 bp in length
            *                    gap of unknown length
            *    95773    99411: contig of 3639 bp in length
            *                    gap of unknown length
            *    99412   103446: contig of 4035 bp in length
            *                    gap of unknown length
            *   103447   108787: contig of 5341 bp in length
            *                    gap of unknown length
            *   108788   116339: contig of 7552 bp in length
            *                    gap of unknown length
            *   116340   124476: contig of 8137 bp in length
            *                    gap of unknown length
            *   124477   133200: contig of 8724 bp in length.
FEATURES             Location/Qualifiers
     source          1. .133200
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-18M17"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    38324 a  27723 c  27998 g  39103 t     52 others
ORIGIN
```

Query Match 53.4%; Score 724.2; DB 43; Length 133200;
Best Local Similarity 89.8%; Pred. No. 1.1e-155;
Matches 799; Conservative 0; Mismatches 89; Indels 2; Gaps 2;

```
Qy     226 cctcagcctcaggcatctgcatctgggaccgacctcctgggctggctgatcaaaga-gga 284
           ||   ||| ||||    ||| | |||||||   | |||||| | |   ||||| || |||
Db   99363 CCGTGGCCCCAGGTTCTTGCTTNTGGGACCCCACCCCTGGGGTTGGCTATCAAGGAGGGA 99304

Qy     285 agcagcagcaatgtctgctgtgggggctgcaactccatacctgcatcatcctggtgatag 344
           |||| ||||| ||| |||||||||||  || |||  ||  |||||||||||||| |||
Db   99303 AGCACCAGCAGTGTTTGCTGTGGGGGCCACAGCTCTGTAT-TGCATCATCCTGGTGGTAG 99245

Qy     345 tcacagtggccgagtgagtttcttgggggcccagcttcctccagaggtggcagcaatggc 404
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   99244 TCACAGTGGCCGAGTGAGTTTTTTGGGGGCCCAGCTTCCTCCAGAGGTGGCAGCAATGGA 99185

Qy     405 ccggctactaggggacctagacaggagcacgttcagaaagttgctgaagtttgtggtcag 464
           || |||||||||||||| || ||||||||| ||| ||||||||||||||||||||||||||
Db   99184 CCAGCTACTAGGGGACTTAAACAGGAGCATGTTTAGAAAGTTGCTGAAGTTTGTGGTCAG 99125

Qy     465 cagcctgcaggggggaggactgccgagacggtgtgcagcgtcttggggtcagcgccaacct 524
           ||||||||| |||||||||||||||| | ||||||| |||||||||||||| |||||||
Db   99124 CAGCCTGCAAGGGGAGGACTGCCGAGAGGCTGTGCAGTGTCTTGGGGTCAGCACCAACCT 99065

Qy     525 gccggaggagcagctgggtgccctgctggcaggcatgcacacactgctccagcaggccct 584
           |||| |||||  |||||||||||| |||||||||||||||| ||||||||||||||||||
Db   99064 GCCGCAGGAGTGGCTGGGTGCCCTACTGGCAGGCATGCACACGCTGCTCCAGCAGGCCCT 99005

Qy     585 ccgtctgccccccaccagcctgaagcctgacaccttcagggaccagctccaggagctctg 644
           || ||| |||||||||||||||||||  |||||||||||||||||||||||| |||||
Db   99004 CCATCTTCCCCCCACCAGCCTGAAGCCCAACACCTTCAGGGACCAGCTCCAGGAACTCTG 98945

Qy     645 catcccccaagacctggtcggggacttggccagcgtggtatttgggagccagcggccccct 704
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   98944 CATCCCCCAAGACCTGGTCGGGGACTTGGCCAGCGTGGTATTTGGGAGCCAGCGGCCCCT 98885
```

```
Qy     705 ccttgattctgtggcccagcagcaggggggcctggctgccgcatgttgctgactttcggtg 764
           |||||||||||||||||||||||||||||||||||||||||| |||||||||||||| |||
Db   98884 CCTTGATTCTGTGGCCCAGCAGCAGGGGGGCCTGGCTGCCCCATGTTGCTGACTTTCAGTG 98825

Qy     765 gcgggtggatgtagcaatctccaccagtgccctggctcgctccctgcagccgagcgtcct 824
           | ||||||||||  |||||||||||||||||||||||||||||| ||||||||||| |||||
Db   98824 GTGGGTGGATGTGACAATCTCCACCAGTGCCCTGGCTCGCTTCCTGCAGCCGAGGGTCCT 98765

Qy     825 gatgcagctgaagctttcagatgggtcagcataccgctttgaggtccccacagccaagtt 884
           |||||||||||||||||||||||| |||||||||||| |||||||||| |||||||||
Db   98764 GATGCAGCTGAAGCTTTCAGATGGGTTAGCATACCGCTTCGAGGTCCCCATAGCCAAGTT 98705

Qy     885 ccaggagctgcggtacagcgtggccctggtcctaaaggagatggcagatctggagaagag 944
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   98704 CCAGGAGCTGCGGTACAGCGTGGCCCTGGTCCTAAAGGAGATGGCAGATCTGGAGAAGAG 98645

Qy     945 gtgtgagcgcagactgcaggactgacccctcacttgaccagtcccattcagatccggctt 1004
           ||||||| |||| |||||||||||||||||||| |||||||||| |||||| || |||||
Db   98644 GTGTGAGGGCAGCCTGCAGGACTGACCCCTCACATGACCAGTCCTATTCAGGTCAGGCTT 98585

Qy    1005 ggacaggcacctgagatggtgccaaagtgcagctgactcttcccacgacagccctggcct 1064
           |||||||||||| |||||||||||||||||||||||||||||||| |||||||||| |||
Db   98584 GGACAGGCACCTCAGATGGTGCCAAAGTGCAGCTGACTCTTCCCAGGACAGCCCTGCCCT 98525

Qy    1065 tcccatcaggcaggctcttcagtgagtgtttgaacgtaattatgtagttt 1114
           |||||| ||||||||||||||   ||    |       ||  |||| |||
Db   98524 TCCCATGAGGCAGGCTCTTCATCCAGGTGCTTTTGTAGATCTTGTATTTT 98475
```

```
RESULT   7
HS79C3R
LOCUS       HS79C3R       347 bp    DNA             PRI       22-OCT-1995
DEFINITION  H.sapiens CpG island DNA genomic MseI fragment, clone 79c3, reverse
            read cpg79c3.rtla.
ACCESSION   Z63154
VERSION     Z63154.1  GI:1035532
KEYWORDS    CpG island; genomic MseI fragment.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 347)
  AUTHORS   MacDonald,M., Huckle,E., Wilkinson,P. and Micklem,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1RQ, England. E-mail contact: humquery@sanger.ac.uk
REFERENCE   2  (bases 1 to 347)
  AUTHORS   Cross,S.H., Charlton,J.A., Nan,X. and Bird,A.P.
  TITLE     Purification of CpG islands using a methylated DNA binding column
  JOURNAL   Nat. Genet. 6 (3), 236-244 (1994)
  MEDLINE   94282070
COMMENT     Vector: pGEM-5Zf(-)
            Clones are available from the UK MRC Human Genome Mapping Project
            Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:
            http://www.hgmp.mrc.ac.uk/ for details
            or contact: biohelp@hgmp.mrc.ac.uk.
FEATURES             Location/Qualifiers
     source          1. .347
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /sex="male"
                     /dev_stage="adult"
                     /tissue_type="blood"
                     /clone_lib="CGI-1"
                     /clone="79c3"
BASE COUNT       65 a    115 c    102 g     54 t     11 others
ORIGIN
```

Query Match 8.2%; Score 110.8; DB 9; Length 347;
Best Local Similarity 83.9%; Pred. No. 2.5e-15;
Matches 146; Conservative 0; Mismatches 26; Indels 2; Gaps 2;

```
Qy    100 acaaccccggaggcgcagaccgtccacacgggacagggacgcccgctcggctccgctccg 159
          |  ||||  ||||||||||  ||||||||||||||||||||||||||||||||  |||||||||
Db    127 AANACCCTGGAGGCGCAGNCCGTCCACACGGGACAGGGACGCCCGCTCGGTTCCGCTCCG 186

Qy    160 cgctgatcctcaaggtctcgcgg-tttccccgccgcccacccggacgccgacg-aaagcc 217
          |||||||||||||||||||||||  |||||||| ||  ||||||||||||| ||| ||||||
Db    187 CGCTGATCCTCAAGGTCTCGCGGTTTTCCCCGTCGNCCACCCGGACGCCNACGNAAAGCC 246

Qy    218 agcgagctcctcagcctcaggcatctgcatctgggaccgacctcctgggctggc 271
          |||||||||||||||||||    ||| |  |  ||||     ||  ||   ||
Db    247 AGCGAGCTCCTCAGCCTCAGGTGACTGGAATTCTGACCCCAGACCACGGGACGC 300
```

```
RESULT   8
AF014882
LOCUS       AF014882     1041 bp    DNA             PRI       06-MAY-1999
DEFINITION  Homo sapiens NADH dehydrogenase subunit 2 (ND2) gene, mitochondrial
            gene encoding mitochondrial protein, partial cds.
ACCESSION   AF014882
VERSION     AF014882.1  GI:2582056
KEYWORDS    .
SOURCE      human.
  ORGANISM  Mitochondrion Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Wise,C.A., Sraml,M. and Easteal,S.
  TITLE     Departure from Neutrality at the Mitochondrial NADH dehydrogenase
            subunit 2 gene in Humans, but not in Chimpanzees
  JOURNAL   Genetics (1997) In press
REFERENCE   2  (bases 1 to 1041)
  AUTHORS   Wise,C.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUL-1997) Human Genetics Group, John Curtin School of
            Medical Research, The Australian National University, Mills Rd,
            Canberra, ACT 0200, Australia
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="Homo sapiens"
                     /organelle="mitochondrion"
                     /strain="Afr1"
                     /db_xref="taxon:9606"
                     /note="Bantu from Durban, South Africa"
     gene            1. .>1041
                     /gene="ND2"
     CDS             1. .>1041
                     /gene="ND2"
                     /codon_start=1
                     /transl_table=2
                     /product="NADH dehydrogenase subunit 2"
                     /protein_id="AAC25441.1"
                     /db_xref="GI:2582057"
                     /translation="MNPLAQPVIYSTIFAGTLITALSSHWFFTWVGLEMNMLAFIPVL
                     TKKMNPRSTEAAIKYFLTQATASMILLMAILFNNMLSGQWTMTNTTNQYSSLMIMMAM
                     AMKLGMAPFHFWVPEVTQGTPLTSGLLLLTWQKLAPISIMYQISPSLNVSLLLTLSIL
                     SIMAGSWGGLNQTQLRKILAYSSITHMGWMMAVLPYNPNMTILNLTIYIILTTTAFLL
                     LNLNSSTTTLLLSRTWNKLTWLTPLIPSTLLSLGGLPPLTGFLPKWAIIEEFTKNNSL
                     IIPTIMATITLLNLYFYLRLIYSTSITLLPMSNNVKMKWQFEHTKPTPFLPTLIALTT
                     LLLPISPFMLMIL"
BASE COUNT      325 a    349 c    100 g    267 t
ORIGIN
```

```
  Query Match             7.7%;  Score 104;  DB 40;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 9e-14;
  Matches  104;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 gcaggcagttgaggtggattaaaccaaacccagctacgcaaaatcttagcatactcctca 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    490 GCAGGCAGTTGAGGTGGATTAAACCAAACCCAGCTACGCAAAATCTTAGCATACTCCTCA 549

Qy     63 attacccacataggatgaataatagcagttctaccgtacaaccc 106
          |||||||||||||||||||||||||||||||||||||||||||||
Db    550 ATTACCCACATAGGATGAATAATAGCAGTTCTACCGTACAACCC 593
```

```
RESULT   9
AF014883
LOCUS       AF014883     1041 bp    DNA             PRI       06-MAY-1999
DEFINITION  Homo sapiens NADH dehydrogenase subunit 2 (ND2) gene, mitochondrial
            gene encoding mitochondrial protein, partial cds.
ACCESSION   AF014883
VERSION     AF014883.1  GI:3287310
KEYWORDS    .
SOURCE      human.
  ORGANISM  Mitochondrion Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Wise,C.A., Sraml,M. and Easteal,S.
  TITLE     Departure from Neutrality at the Mitochondrial NADH dehydrogenase
            subunit 2 gene in Humans, but not in Chimpanzees
  JOURNAL   Genetics (1997) In press
REFERENCE   2  (bases 1 to 1041)
  AUTHORS   Wise,C.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUL-1997) Human Genetics Group, John Curtin School of
            Medical Research, The Australian National University, Mills Rd,
            Canberra, ACT 0200, Australia
COMMENT     On Jul 2, 1998 this sequence version replaced gi:2582058.
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="Homo sapiens"
                     /organelle="mitochondrion"
                     /strain="Afr2"
                     /db_xref="taxon:9606"
                     /note="Bantu from Durban, South Africa"
     gene            1. .>1041
                     /gene="ND2"
     CDS             1. .>1041
                     /gene="ND2"
                     /codon_start=1
                     /transl_table=2
                     /product="NADH dehydrogenase subunit 2"
                     /protein_id="AAC25442.1"
                     /db_xref="GI:3287311"
                     /translation="MNPLAQPVIYSTIFAGTLITALSSHWFFTWVGLEMNMLAFIPVL
                     TKKMNPRSTEAAIKYFLTQATASMILLMAILFNNMLSGQWTMTNTTNQYSSLMIMMAM
                     AMKLGMAPFHFWVPEVTQGTPLTSGLLLLTWQKLAPISIMYQISPSLNVSLLLTLSIL
                     SIMAGSWGGLNQTQLRKILAYSSITHMGWMMAVLPYNPNMTILNLTIYIILTTTAFLL
                     LNLNSSTTTLLLSRTWNKLTWLTPLIPSTLLSLGGLPPLTGFLPKWAIIEEFTKNNSL
                     IIPTIMATITLLNLYFYLRLIYSTSITLLPMSNNVKMKWQFEHTKPTPLLPTLIALTT
                     LLLPISPFMLMIL"
BASE COUNT      325 a    350 c    100 g    266 t
ORIGIN
```

```
  Query Match             7.7%;  Score 104;  DB 40;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 9e-14;
  Matches  104;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 gcaggcagttgaggtggattaaaccaaacccagctacgcaaaatcttagcatactcctca 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    490 GCAGGCAGTTGAGGTGGATTAAACCAAACCCAGCTACGCAAAATCTTAGCATACTCCTCA 549

Qy     63 attacccacataggatgaataatagcagttctaccgtacaaccc 106
          |||||||||||||||||||||||||||||||||||||||||||||
Db    550 ATTACCCACATAGGATGAATAATAGCAGTTCTACCGTACAACCC 593
```

```
RESULT  10
AF014884
LOCUS       AF014884     1041 bp    DNA             PRI       06-MAY-1999
DEFINITION  Homo sapiens NADH dehydrogenase subunit 2 (ND2) gene, mitochondrial
            gene encoding mitochondrial protein, partial cds.
```

```
ACCESSION   AF014884
VERSION     AF014884.2  GI:4755089
KEYWORDS    .
SOURCE      human.
  ORGANISM  Mitochondrion Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Wise,C.A., Sraml,M. and Easteal,S.
  TITLE     Departure from Neutrality at the Mitochondrial NADH dehydrogenase
            subunit 2 gene in Humans, but not in Chimpanzees
  JOURNAL   Genetics (1997) In press
REFERENCE   2  (bases 1 to 1041)
  AUTHORS   Wise,C.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUL-1997) Human Genetics Group, John Curtin School of
            Medical Research, The Australian National University, Mills Rd,
            Canberra, ACT 0200, Australia
COMMENT     On May 6, 1999 this sequence version replaced gi:3287312.
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="Homo sapiens"
                     /organelle="mitochondrion"
                     /strain="Afr3"
                     /db_xref="taxon:9606"
                     /note="Bantu from Durban, South Africa"
     gene            1. .>1041
                     /gene="ND2"
     CDS             1. .>1041
                     /gene="ND2"
                     /codon_start=1
                     /transl_table=2
                     /product="NADH dehydrogenase subunit 2"
                     /protein_id="AAC25443.1"
                     /db_xref="GI:2582061"
                     /translation="MNPLAQPVIYSTIFAGTLITALSSHWFFTWVGLEMNMLAFIPVL
                     TKKMNPRSTEAAIKYFLTQATASMILLMAILFNNMLSGQWTMTNTTNQYSSLMIMMAM
                     AMKLGMAPFHFWVPEVTQGTPLTSGLLLLTWQKLAPISIMYQISPSLNVSLLLTLSIL
                     SIMAGSWGGLNQTQLRKILAYSSITHMGWMMAVLPYNPNMTILNLTIYIILTTTAFLL
                     LNLNSSTTTLLLSRTWNKLTWLTPLIPSTLLSLGGLPPLTGFLPKWAIIEEFTKNNSL
                     IIPTIMATITLLNLYFYLRLIYSTSITLLPMSNNVKMKWQFEHTKPTPFLPTLIALTT
                     LLLPISPFMLMIL"
BASE COUNT      325 a    349 c    100 g    267 t
ORIGIN
```

Query Match 7.7%; Score 104; DB 40; Length 1041;
Best Local Similarity 100.0%; Pred. No. 9e-14;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      3 gcaggcagttgaggtggattaaaccaaacccagctacgcaaaatcttagcatactcctca 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    490 GCAGGCAGTTGAGGTGGATTAAACCAAACCCAGCTACGCAAAATCTTAGCATACTCCTCA 549

Qy     63 attacccacataggatgaataatagcagttctaccgtacaaccc 106
          |||||||||||||||||||||||||||||||||||||||||||||
Db    550 ATTACCCACATAGGATGAATAATAGCAGTTCTACCGTACAACCC 593
```

```
RESULT  11
AF014885
LOCUS       AF014885     1041 bp    DNA             PRI       06-MAY-1999
DEFINITION  Homo sapiens NADH dehydrogenase subunit 2 (ND2) gene, mitochondrial
            gene encoding mitochondrial protein, partial cds.
ACCESSION   AF014885
VERSION     AF014885.1  GI:2582062
KEYWORDS    .
SOURCE      human.
  ORGANISM  Mitochondrion Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Wise,C.A., Sraml,M. and Easteal,S.
  TITLE     Departure from Neutrality at the Mitochondrial NADH dehydrogenase
            subunit 2 gene in Humans, but not in Chimpanzees
  JOURNAL   Genetics (1997) In press
REFERENCE   2  (bases 1 to 1041)
  AUTHORS   Wise,C.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUL-1997) Human Genetics Group, John Curtin School of
            Medical Research, The Australian National University, Mills Rd,
            Canberra, ACT 0200, Australia
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="Homo sapiens"
                     /organelle="mitochondrion"
                     /strain="Afr4"
                     /db_xref="taxon:9606"
                     /note="Bantu from Durban, South Africa"
     gene            1. .>1041
                     /gene="ND2"
     CDS             1. .>1041
                     /gene="ND2"
                     /codon_start=1
                     /transl_table=2
                     /product="NADH dehydrogenase subunit 2"
                     /protein_id="AAC25444.1"
                     /db_xref="GI:2582063"
                     /translation="MNPLAQPVIYSTIFAGTLITALSSHWFFTWVGLEMNMLAFIPVL
                     TKKMNPRSTEAAIKYFLTQATASMILLMAILFNNMLSGQWTMTNTTNQYSSLMIMMAM
                     AMKLGMAPFHFWVPEVTQGTPLTSGLLLLTWQKLAPISIMYQISPSLNVSLLLTLSIL
                     SIMAGSWGGLNQTQLRKILAYSSITHMGWMMAVLPYNPNMTILNLTIYIILTTTAFLL
                     LNLNSSTTTLLLSRTWNKLTWLTPLIPSTLLSLGGLPPLTGFLPKWAIIEEFTKNNSL
                     IIPTIMATITLLNLYFYLRLIYSTSITLLPMSNNVKMKWQFEHTKPTPFLPTLIALTT
                     LLLPISPFMLMIL"
BASE COUNT      325 a    349 c    100 g    267 t
ORIGIN
```

Query Match 7.7%; Score 104; DB 40; Length 1041;
Best Local Similarity 100.0%; Pred. No. 9e-14;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      3 gcaggcagttgaggtggattaaaccaaacccagctacgcaaaatcttagcatactcctca 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    490 GCAGGCAGTTGAGGTGGATTAAACCAAACCCAGCTACGCAAAATCTTAGCATACTCCTCA 549

Qy     63 attacccacataggatgaataatagcagttctaccgtacaaccc 106
          |||||||||||||||||||||||||||||||||||||||||||||
Db    550 ATTACCCACATAGGATGAATAATAGCAGTTCTACCGTACAACCC 593
```

```
RESULT  12
AF014886
LOCUS       AF014886     1041 bp    DNA             PRI       06-MAY-1999
DEFINITION  Homo sapiens NADH dehydrogenase subunit 2 (ND2) gene, mitochondrial
            gene encoding mitochondrial protein, partial cds.
ACCESSION   AF014886
VERSION     AF014886.1  GI:3287313
KEYWORDS    .
SOURCE      human.
  ORGANISM  Mitochondrion Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Wise,C.A., Sraml,M. and Easteal,S.
  TITLE     Departure from Neutrality at the Mitochondrial NADH dehydrogenase
            subunit 2 gene in Humans, but not in Chimpanzees
  JOURNAL   Genetics (1997) In press
REFERENCE   2  (bases 1 to 1041)
  AUTHORS   Wise,C.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUL-1997) Human Genetics Group, John Curtin School of
            Medical Research, The Australian National University, Mills Rd,
            Canberra, ACT 0200, Australia
COMMENT     On Jul 2, 1998 this sequence version replaced gi:2582064.
```

```
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="Homo sapiens"
                     /organelle="mitochondrion"
                     /strain="Afr5"
                     /db_xref="taxon:9606"
                     /note="Bantu from Durban, South Africa"
     gene            1. .>1041
                     /gene="ND2"
     CDS             1. .>1041
                     /gene="ND2"
                     /codon_start=1
                     /transl_table=2
                     /product="NADH dehydrogenase subunit 2"
                     /protein_id="AAC25445.1"
                     /db_xref="GI:3287314"
                     /translation="MNPLAQPVIYSTIFAGTLITALSSHWFFTWVGLEMNMLAFIPVL
                     TKKMNPRSTEAAIKYFLTQATASMILLMAILFNNMLSGQWTMTNTTNQYSSLMIMMAM
                     AMKLGMAPFHFWVPEVTQGTPLTSGLLLLTWQKLAPISIMYQISPSLNVSLLLTLSIL
                     SIMAGSWGGLNQTQLRKILAYSSITHMGWMMAVLPYNPNMTILNLTIYIILTTTaFLL
                     LNLNSSTTTLLLSRTWNKLTWLTPLIPSTLLSLGGLPPLTGFLPKWAIIEEFTKNNSL
                     IIPTIMATITLLNLYFYLRLIYSTSITLLPMSNNVKMKWQFEHTKPTPLLPTLIALTT
                     LLLPISPFMLMIL"
BASE COUNT       325 a    350 c    100 g    266 t
ORIGIN


  Query Match             7.7%;  Score 104;  DB 40;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 9e-14;
  Matches  104;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 gcaggcagttgaggtggattaaaccaaacccagctacgcaaaatcttagcatactcctca 62
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     490 GCAGGCAGTTGAGGTGGATTAAACCAAACCCAGCTACGCAAAATCTTAGCATACTCCTCA 549

Qy      63 attacccacataggatgaataatagcagttctaccgtacaaccc 106
           |||||||||||||||||||||||||||||||||||||||||||
Db     550 ATTACCCACATAGGATGAATAATAGCAGTTCTACCGTACAACCC 593


RESULT  13
AF014887
LOCUS       AF014887     1041 bp    DNA             PRI       06-MAY-1999
DEFINITION  Homo sapiens NADH dehydrogenase subunit 2 (ND2) gene, mitochondrial
            gene encoding mitochondrial protein, partial cds.
ACCESSION   AF014887
VERSION     AF014887.1  GI:2582066
KEYWORDS    .
SOURCE      human.
  ORGANISM  Mitochondrion Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Wise,C.A., Sraml,M. and Easteal,S.
  TITLE     Departure from Neutrality at the Mitochondrial NADH dehydrogenase
            subunit 2 gene in Humans, but not in Chimpanzees
  JOURNAL   Genetics (1997) In press
REFERENCE   2  (bases 1 to 1041)
  AUTHORS   Wise,C.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUL-1997) Human Genetics Group, John Curtin School of
            Medical Research, The Australian National University, Mills Rd,
            Canberra, ACT 0200, Australia
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="Homo sapiens"
                     /organelle="mitochondrion"
                     /strain="Eur1"
                     /db_xref="taxon:9606"
                     /note="European: Anglo Celt from Canberra, Australia"
     gene            1. .>1041
                     /gene="ND2"
     CDS             1. .>1041
                     /gene="ND2"
                     /codon_start=1
                     /transl_table=2
                     /product="NADH dehydrogenase subunit 2"
                     /protein_id="AAC25446.1"
                     /db_xref="GI:2582067"
                     /translation="MNPLAQPVIYSTIFAGTLITALSSHWFFTWVGLEMNMLAFIPVL
                     TKKMNPRSTEAAIKYFLTQATASMILLMAILFNNMLSGQWTMTNTTNQYSSLMIMMAM
                     AMKLGMAPFHFWVPEVTQGTPLTSGLLLLTWQKLAPISIMYQISPSLNVSLLLTLSIL
                     SIMAGSWGGLNQTQLRKILAYSSITHMGWMMAVLPYNPNMTILNLTIYIILTTTAFLL
                     LNLNSSTTTLLLSRTWNKLTWLTPLIPSTLLSLGGLPPLTGFLPKWAIIEEFTKNNSL
                     IIPTIMATITLLNLYFYLRLIYSTSITLLPMSNNVKMKWQFEHTKPTPFLPTLIALTT
                     LLLPISPFMLMIL"
BASE COUNT       325 a    349 c    100 g    267 t
ORIGIN


  Query Match             7.7%;  Score 104;  DB 40;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 9e-14;
  Matches  104;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 gcaggcagttgaggtggattaaaccaaacccagctacgcaaaatcttagcatactcctca 62
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     490 GCAGGCAGTTGAGGTGGATTAAACCAAACCCAGCTACGCAAAATCTTAGCATACTCCTCA 549

Qy      63 attacccacataggatgaataatagcagttctaccgtacaaccc 106
           |||||||||||||||||||||||||||||||||||||||||||
Db     550 ATTACCCACATAGGATGAATAATAGCAGTTCTACCGTACAACCC 593


RESULT  14
AF014888
LOCUS       AF014888     1041 bp    DNA             PRI       06-MAY-1999
DEFINITION  Homo sapiens NADH dehydrogenase subunit 2 (ND2) gene, mitochondrial
            gene encoding mitochondrial protein, partial cds.
ACCESSION   AF014888
VERSION     AF014888.1  GI:3287315
KEYWORDS    .
SOURCE      human.
  ORGANISM  Mitochondrion Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Wise,C.A., Sraml,M. and Easteal,S.
  TITLE     Departure from Neutrality at the Mitochondrial NADH dehydrogenase
            subunit 2 gene in Humans, but not in Chimpanzees
  JOURNAL   Genetics (1997) In press
REFERENCE   2  (bases 1 to 1041)
  AUTHORS   Wise,C.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUL-1997) Human Genetics Group, John Curtin School of
            Medical Research, The Australian National University, Mills Rd,
            Canberra, ACT 0200, Australia
COMMENT     On Jul 2, 1998 this sequence version replaced gi:2582068.
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="Homo sapiens"
                     /organelle="mitochondrion"
                     /strain="Eur2"
                     /db_xref="taxon:9606"
                     /note="European: Anglo Celt from Canberra, Australia"
     gene            1. .>1041
                     /gene="ND2"
     CDS             1. .>1041
                     /gene="ND2"
                     /codon_start=1
                     /transl_table=2
                     /product="NADH dehydrogenase subunit 2"
                     /protein_id="AAC25447.1"
                     /db_xref="GI:3287316"
                     /translation="MNPLAQPVIYSTIFAGTLITALSSHWFFTWVGLEMNMLAFIPVL
                     TKKMNPRSTEAAIKYFLTQATASMILLMAILFNNMLSGQWTMTNTTNQYSSLMIMMAM
                     AMKLGMAPFHFWVPEVTQGTPLTSGLLLLTWQKLAPISIMYQISPSLDVSLLLTLSIL
```

```
                     SIMAGSWGGLNQTQLRKILAYSSITHMGWMMAVLPYNPNMTILNLTIYIILTTTAFLL
                     LNLNSSTTTLLLSRTWNKLTWLTPLIPSTLLSLGGLPPLTGFLPKWAIIEEFTKNNSL
                     IIPTIMATITLLNLYFYLRLIYSTSITLLPMSNNVKMKWQFEHTKPTPFLPTLIALTT
                     LLLPISPFMLMIL"
BASE COUNT      324 a    349 c    101 g    267 t
ORIGIN


  Query Match             7.7%;  Score 104;  DB 40;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 9e-14;
  Matches  104;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 gcaggcagttgaggtggattaaaccaaaacccagctacgcaaaatcttagcatactcctca 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    490 GCAGGCAGTTGAGGTGGATTAAACCAAACCCAGCTACGCAAAATCTTAGCATACTCCTCA 549

Qy     63 attacccacataggatgaataatagcagttctaccgtacaaccc 106
          |||||||||||||||||||||||||||||||||||||||||||||
Db    550 ATTACCCACATAGGATGAATAATAGCAGTTCTACCGTACAACCC 593


RESULT  15
AF014889
LOCUS       AF014889     1041 bp    DNA             PRI       06-MAY-1999
DEFINITION  Homo sapiens NADH dehydrogenase subunit 2 (ND2) gene, mitochondrial
            gene encoding mitochondrial protein, partial cds.
ACCESSION   AF014889
VERSION     AF014889.1  GI:3287317
KEYWORDS    .
SOURCE      human.
  ORGANISM  Mitochondrion Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Wise,C.A., Sraml,M. and Easteal,S.
  TITLE     Departure from Neutrality at the Mitochondrial NADH dehydrogenase
            subunit 2 gene in Humans, but not in Chimpanzees
  JOURNAL   Genetics (1997) In press
REFERENCE   2  (bases 1 to 1041)
  AUTHORS   Wise,C.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUL-1997) Human Genetics Group, John Curtin School of
            Medical Research, The Australian National University, Mills Rd,
            Canberra, ACT 0200, Australia
COMMENT     On Jul 2, 1998 this sequence version replaced gi:2582070.
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="Homo sapiens"
                     /organelle="mitochondrion"
                     /strain="Eur3"
                     /db_xref="taxon:9606"
                     /note="European: Anglo Celt from Canberra, Australia"
     gene            1. .>1041
                     /gene="ND2"
     CDS             1. .>1041
                     /gene="ND2"
                     /codon_start=1
                     /transl_table=2
                     /product="NADH dehydrogenase subunit 2"
                     /protein_id="AAC25448.1"
                     /db_xref="GI:3287318"
                     /translation="MNPLAQPVIYSTIFAGTLITALSSHWFFTWVGLEMNMLAFIPVL
                     TKKMNPRSTEAAIKYFLTQATASMILLMAILFNNMLSGQWTMTSTTNQYSSLMIMMAM
                     AMKLGMAPFHFWVPEVTQGTPLTSGLLLLTWQKLAPISIMYQISPSLNVSLLLTLSIL
                     SIMAGSWGGLNQTQLRKILAYSSITHMGWMMAVLPYNPNMTILNLTIYIILTTTAFLL
                     LNLNSSTTTLLLSRTWNKLTWLTPLIPSTLLSLGGLPPLTGFLPKWAIIEEFTKNNSL
                     IIPTIMATITLLNLYFYLRLIYSTSITLLPMSNNVKMKWQFEHTKPTPFLPTLIALTT
                     LLLPISPFMLMIL"
BASE COUNT      324 a    349 c    101 g    267 t
ORIGIN


  Query Match             7.7%;  Score 104;  DB 40;  Length 1041;
```

```
  Best Local Similarity   100.0%;  Pred. No. 9e-14;
  Matches  104;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 gcaggcagttgaggtggattaaaccaaaacccagctacgcaaaatcttagcatactcctca 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    490 GCAGGCAGTTGAGGTGGATTAAACCAAACCCAGCTACGCAAAATCTTAGCATACTCCTCA 549

Qy     63 attacccacataggatgaataatagcagttctaccgtacaaccc 106
          |||||||||||||||||||||||||||||||||||||||||||||
Db    550 ATTACCCACATAGGATGAATAATAGCAGTTCTACCGTACAACCC 593


Search completed: May 15, 2000, 02:00:07
Job time: 11526 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: May 15, 2000, 01:45:56 ; Search time 84.19 Seconds
(without alignments)
4026.729 Million cell updates/sec

Title: US-09-223-796-3
Perfect score: 1355
Sequence: 1 gggcaggcagttgaggtgga.....gtgtttcaggcagggcccgg 1355

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	469.6	34.7	1100	1	V18890	Rat Hypertension r
	2	338.8	25.0	701	1	V18891	Human Hypertension
	3	175.8	13.0	186	1	T25721	Human gene signatu
	4	104.4	7.7	448	1	V86703	EST clone BG461. N
	5	45.2	3.3	350	1	V38226	Human mucosal adre
c	6	44	3.2	1771	1	T62138	Leishmania tropica
c	7	44	3.2	1771	1	V47558	Leishmania antigen
	8	41.8	3.1	1536	1	V38201	Human mucosal adre
	9	41.6	3.1	1488	1	V38202	Human mucosal adre
	10	41.6	3.1	1539	1	T39751	Human mucosal adr
	11	41.6	3.1	1539	1	X35280	cDNA encoding huma
	12	41.6	3.1	1624	1	T39750	Human mucosal addr
	13	41.6	3.1	1624	1	X35279	cDNA encoding huma
	14	40.2	3.0	1338	1	X17775	HTLAR33 encoding c
	15	40.2	3.0	2394	1	X17774	YAK-1 related seri
	16	39.6	2.9	1879	1	Q11126	Sequence encoding
	17	39.6	2.9	5267	1	V32449	Human receptor tyr
	18	39	2.9	1297	1	V84341	Human neuro-growth
	19	38.6	2.8	1503	1	V72539	Thermus flavus amy
c	20	38.6	2.8	5173	1	T89783	Human Immunodefici
c	21	38.4	2.8	2000	1	N71065	Gene encoding Plas
	22	38.4	2.8	49272	1	V35000	Mycobacteriophage
	23	37.8	2.8	2417	1	Q56106	DNA encoding a lip
	24	37.8	2.8	32207	1	V73805	KSHV LUR DNA (nucl
	25	37.8	2.8	137507	1	V19941	KSHV long unique c
	26	37.4	2.8	3435	1	T35869	Human DNA polymera
c	27	37	2.7	1176	1	V23491	Pseudomonas OrfY s
c	28	37	2.7	17612	1	V23494	Pseudomonas Xpc, O
	29	36.6	2.7	888	1	T59269	Streptomyces prist
	30	36.6	2.7	2821	1	T90730	Human KVLQT1 full-
	31	36.6	2.7	2821	1	T94004	DNA encoding human
c	32	36.6	2.7	2888	1	T59268	Streptomyces prist
	33	36.4	2.7	1724	1	X32493	Seq ID No:8' used i
c	34	35.8	2.6	541	1	T76907	S. glaucescens Hst
	35	35.8	2.6	630	1	T85894	Malassezia fungus
	36	35.8	2.6	864	1	Q98748	cDNA for human pro
	37	35.8	2.6	921	1	Q98749	Full length cDNA f
	38	35.8	2.6	1712	1	T61835	Pseudomonas aerugi
c	39	35.6	2.6	432	1	Q43031	Collagen-like poly
c	40	35.6	2.6	432	1	T16765	Collagen-like poly
	41	35.6	2.6	595	1	Q50041	ICP34.5 fragment.
c	42	35.6	2.6	756	1	Q43032	Collagen-like poly
c	43	35.6	2.6	756	1	T16766	Collagen-like poly
	44	35.6	2.6	1335	1	Q23295	HSV-1 (F) ICP34.5
	45	35.4	2.6	432	1	Q43031	Collagen-like poly

ALIGNMENTS

RESULT 1
V18890
ID V18890 standard; cDNA; 1100 BP.
AC V18890;
DT 09-JUN-1998 (first entry)
DE Rat Hypertension related calcium regulated gene (HCaRG) cDNA.
KW Hypertension related calcium regulated gene; HCaRG; rat parathyroid;
KW extracellular calcium concentration; antibody; hypertension; ss;
KW hyperthyroidism; osteoporosis; heart failure; diabetes; stroke;
KW cancer; inflammatory disease; asthma.
OS Rattus rattus.
FH Key Location/Qualifiers
FT CDS 131. .806
FT /*tag= a
FT /product= "Hypertension related calcium regulator"
FT misc_feature 173. .196
FT /*tag= b
FT /note= "EF-hand like motif"
PN WO9749807-A2.
PD 31-DEC-1997.
PF 23-JUN-1997; CA0439.
PR 21-JUN-1996; US-667495.
PA (GOSS/) GOSSARD F.
PA (HAME/) HAMET P.
PA (LEWA/) LEWANCZUK R.
PA (TREM/) TREMBLAY J.
PI Gossard F, Hamet P, Lewanczuk R, Tremblay J;
DR WPI; 98-077171/07.
DR P-PSDB; W37723.
PT Hypertension related calcium regulated gene - useful to develop
PT products to treat or detect, e.g. hypertension, stroke,
PT osteoporosis, heart failure, cancer, diabetes or asthma
PS Claim 2; Pages 25-26; 46pp; English.
CC This nucleic acid sequence encodes the hypertension related calcium
CC regulated gene (HCaRG), which was isolated from the rat parathyroid.
CC Its expression is regulated by extracellular calcium concentration.
CC An antibody against the protein, can be used to detect or modulate
CC (e.g. enhance or inhibit) abnormal calcium levels. They can
CC specifically be used to detect or treat, e.g. hypertension,
CC hyperthyroidism, osteoporosis, heart failure, diabetes, stroke, cancer,
CC inflammatory disease, and asthma.
SQ Sequence 1100 BP; 260 A; 279 C; 315 G; 246 T;

Query Match 34.7%; Score 469.6; DB 1; Length 1100;
Best Local Similarity 77.2%; Pred. No. 4.2e-114;
Matches 571; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

```
Qy   247 tctgggaccgacctcctgggctggctgatcaaagaggaagcagcagcaatgtctgctgtg 306
         | |||    ||||| ||   | | ||| ||||  |    | || |   |||||||| ||
Db    84 TGTGGTTAGGACCTTCTCTTCGGACTGGTCAAGAAACGGGAAGAAAGGATGTCTGCTTTG 143

Qy   307 ggggctgcaactccatacctgcatcatcctggtgatagtcacagtggccgagtgagtttc 366
         |||||||| |||||||| |||| ||||| | ||| ||||||||||||||| || ||||||
Db   144 GGGGCTGCAGCTCCATACTTGCACCATCCCGCTGACAGTCACAGTGGCCGGGTCAGTTTC 203

Qy   367 ttgggggcccagcttcctccagaggtggcagcaatggcccggctactaggggacctagac 426
```

```
Db    204 CTGGGTTCCCAGCCCTCTCCAGAAGTGACGGCCGTGGCTCAGCTCTTGAAGGACTTAGAC 263

Qy    427 aggagcacgttcagaaagttgctgaagtttgtggtcagcagcctgcagggggaggactgc 486

Db    264 AGGAGCACCTTCAGAAAGTTGTTGAAACTTGTAGTCGGGGCCCTGCATGGGAAAGACTGC 323

Qy    487 cgagacggtgtgcagcgtcttggggtcagcgccaacctgccggaggagcagctgggtgcc 546

Db    324 AGAGAAGCTGTGGAGCAACTTGGTGCCAGCGCCAACCTGTCAGAAGAGCGTCTGGCCGTC 383

Qy    547 ctgctggcaggcatgcacacactgctccagcaggccctccgtctgccccccaccagcctg 606

Db    384 CTGCTGGCGGGCACACACACCCTGCTCCAGCAGGCTCTCCGGCTGCCCCCTGCTAGTCTA 443

Qy    607 aagcctgacaccttcagggaccagctccaggagctctgcatcccccaagacctggtcggg 666

Db    444 AAGCCAGATGCCTTCCAGGAAGAGCTCCAGGAACTTGGCATTCCTCAGGATCTAATTGGA 503

Qy    667 gacttggccagcgtggtatttgggagccagcggcccctccttgattctgtggcccagcag 726

Db    504 GATTTGGCCAGTTTGGCATTTGGGAGTCAACGCCCTCTTCTCGACTCTGTAGCCCAACAG 563

Qy    727 caggggggcctggctgccgcatgttgctgactttcggtggcgggtggatgtagcaatctcc 786

Db    564 CAGGGATCCTCGCTGCCTCACGTGTCTTACTTCCGGTGGCGGGTGGACGTGGCCATCTCA 623

Qy    787 accagtgccctggctcgctccctgcagccgagcgtcctgatgcagctgaagctttcagat 846

Db    624 ACCAGCGCTCAGTCCCGCTCCCTGCAACCGAGTGTTCTCATGCAGCTGAAGCTCACAGAT 683

Qy    847 gggtcagcataccgctttgaggtccccacagccaagttccaggagctgcggtacagcgtg 906

Db    684 GGATCTGCACACCGCTTCGAGGTGCCCATAGCCAAATTTCAGGAGCTGCGGTACAGTGTA 743

Qy    907 gccctggtcctaaaggagatggcagatctggagaagaggtgtgagcgcagactgcaggac 966

Db    744 GCCTTGGTCCTTAAGGAGATGGCAGAACTGGAGAAGAAGTGTGAGCGCAAACTGCAGGAC 803

Qy    967 tgacccctcacttgaccagt 986

Db    804 TGACTGAACCCTGGTACTGT 823
```

RESULT 2
V18891
ID V18891 standard; cDNA; 701 BP.
AC V18891;
DT 09-JUN-1998 (first entry)
DE Human Hypertension related calcium regulated gene (HCaRG) cDNA.
KW Hypertension related calcium regulated gene; HCaRG; human parathyroid;
KW extracellular calcium concentration; antibody; hypertension; ss;
KW hyperthyroidism; osteoporosis; heart failure; diabetes; stroke;
KW cancer; inflammatory disease; asthma.
OS Homo sapiens.
FH Key Location/Qualifiers
FT misc_feature 224. .245
FT /*tag= a
FT /note= "EF-hand like motif"
PN WO9749807-A2.
PD 31-DEC-1997.
PF 23-JUN-1997; CA0439.
PR 21-JUN-1996; US-667495.
PA (GOSS/) GOSSARD F.
PA (HAME/) HAMET P.
PA (LEWA/) LEWANCZUK R.
PA (TREM/) TREMBLAY J.
PI Gossard F, Hamet P, Lewanczuk R, Tremblay J;
DR WPI; 98-077171/07.
PT Hypertension related calcium regulated gene - useful to develop
PT products to treat or detect, e.g. hypertension, stroke,
PT osteoporosis, heart failure, cancer, diabetes or asthma
PS Claim 3; Pages 27-28; 46pp; English.

CC This nucleic acid sequence encodes the hypertension related calcium
CC regulated gene (HCaRG), which was isolated from the human parathyroid.
CC Its expression is regulated by extracellular calcium concentration.
CC The sequence displays 75 percent homology to the rat form, with
CC conserved areas such as the EF-hand like motifs and the initiating
CC codon. An antibody against the protein, can be used to detect or
CC modulate (e.g. enhance or inhibit) abnormal calcium levels. They can
CC specifically be used to detect or treat, e.g. hypertension,
CC hyperthyroidism, osteoporosis, heart failure, diabetes, stroke, cancer,
CC inflammatory disease, and asthma.
SQ Sequence 701 BP; 137 A; 204 C; 217 G; 139 T;

Query Match 25.0%; Score 338.8; DB 1; Length 701;
Best Local Similarity 72.2%; Pred. No. 8e-80;
Matches 659; Conservative 0; Mismatches 30; Indels 224; Gaps 8;

```
Qy    100 acaaccccggaggcgcagaccgtccacacgggacagggacgcccgctcggctccgctccg 159

Db      8 AAAACCCTGGAGGCGCAGACCGTCCACACGAGACAGGGACGCCCGCTCGGCTCCGCTCCG 67

Qy    160 cgctgatcctcaaggtctcgcggtttccccgccgcccacccggacgccgacgaaagccag 219

Db     68 -GCTGATCCTCAAGGTCTCGCGGTTTCCCCNCCGCCCACCCGGACGCCGACGAAAGCCAG 126

Qy    220 cgagctcctcagcctcaggcagctctgcatctgggaccgacctcctgggctggctgatcaaa 279

Db    127 CGAGCTCCTCAGTCT----------CGACCTCCTGGGCTGGCTGATCAAA 166

Qy    280 gaggaagcagcagcaatgtctgctgtgggggctgcaactccatacctgcatcatcctggt 339

Db    167 GAGGAAGCAGCAGCAATGTCTGCTTTGGGGACTGCAACTCCATACCTGCATCATCCTGGT 226

Qy    340 gatagtcacagtggccgagtgagtttcttgggggcccagcttcctccagaggtggcagca 399

Db    227 GATAGTCACAGTGGCCGAGTGAGTTTCTTGGGGGCCCAGCTTCCTCCAGAGGTGGCAGCA 286

Qy    400 atggcccggctactaggggacctagacaggagcacgttcagaaagttgc-tgaagtttgt 458

Db    287 ATGGCCCAGCTACTAGGGGACCTAGACAGGAGCACGTTCAGAAAGTTGCTTGAAGTTTGT 346

Qy    459 ggtcagcagcctgcaggggggaggactgccgagacggtgtgcagcgtcttggggtcagcgc 518

Db    347 GGTCAGCAGCCTGCAGGGGGAGGATTGCCGAGAGTGTTGCAGCGTNTTTGGGGTCAGCGC 406

Qy    519 caacctgccggaggagcagctgggtgccctgctggcaggcatgcacacactgctccagca 578

Db    407 CAACCTGCCGGAGGAGCAGTTNGGTTGCCT---------- 437

Qy    579 ggccctccgtctgccccccaccagcctgaagcctgacaccttcagggaccagctccagga 638

Db    437 ---------- 437

Qy    639 gctctgcatcccccaagacctggtcggggacttggccagcgtggtatttgggagccagcg 698

Db    437 ---------- 437

Qy    699 gcccctccttgattctgtggcccagcagcaggggggcctggctgccgcatgttgctgactt 758

Db    437 ----------TGTTGGCAGGTT 448

Qy    759 tcggtggcgggtggatgtag-caatctccacc-agtgccctggctcgctccctgcagccg 816

Db    449 TCGGTGGCGGGTGGATGTAGCCAATCTCCACCNAGTGCCCTGGCTCGCTCCCTGCAGCCG 508

Qy    817 agcgtcctgatgcagctg-aagctttcagatgggtcagcata-ccgctttgaggtcccca 874

Db    509 AGCGTCCTGATGCAGCTGTAAGCTTTCAGATGGGTCAGCATACCCGCTTTGAGGTCCCCA 568

Qy    875 cagccaagttccaggagctgcggtacagcgtggccctggtcctaaaggagatggcagatc 934

Db    569 CAGCCAAGTTCCAGGAGCTGCGGTACAGCGTGGCCCTGGTCCTAAAGGAGATGGCAGATC 628
```

```
Qy    935 tggagaagaggtgtgagcgcagactgcaggactgacccctcacttgaccagtcccattca 994
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    629 TGGAGAAGAGGTGTGAGCGCAGACTGCAGGACTGACCCCTCACTTGACCAGTCCCATTCA 688

Qy    995 gatccggcttgga 1007
          |||||||||||||
Db    689 GATCCGGCTTGGA 701
```

RESULT 3
T25721
ID T25721 standard; cDNA to mRNA; 186 BP.
AC T25721;
DT 10-OCT-1996 (first entry)
DE Human gene signature HUMGS07932.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN WO9514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI; 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 1914; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 186 BP; 46 A; 40 C; 43 G; 54 T;

Query Match 13.0%; Score 175.8; DB 1; Length 186;
Best Local Similarity 97.8%; Pred. No. 3e-37;
Matches 177; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy    995 gatccggcttggacaggcacctgagatggtgccaaagtgcagctgactcttcccacgaca 1054
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GATCCGGCTTGGACAGGCACCTGAGATGGTGCCAAAGTGCAGCTGACTCTTCCCACGACA 60

Qy   1055 gccctggccttcccatcaggcaggctcttcagtgagtgtttgaacgtaattatgtagttt 1114
          |||||| |||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     61 GCCCTGCCCTTCCCATGAGGCAGGCTCTTCAGTGAGTGTTTGAACGTAATTATGTAGTTT 120

Qy   1115 tctgtttaattgaaaaagagagctatgccttttttttcttttttggaagtaaagcagctaaa 1174
          |||||||||||||||||| |||||||||||||||||||||||||| ||||||||||||||
Db    121 TCTGTTTAATTGAAAAAGNGAGCTATGCCTTTTTTTCTTTTTGGANGTAAAGCAGCTAAA 180

Qy   1175 a 1175
          |
Db    181 A 181
```

RESULT 4
V86703
ID V86703 standard; cDNA; 448 BP.
AC V86703;
DT 27-APR-1999 (first entry)
DE EST clone BG461.
KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
OS Homo sapiens.
PN WO9845435-A2.
PD 15-OCT-1998.
PF 10-APR-1998; U06954.
PR 10-APR-1997; US-835913.
PA (GEMY) GENETICS INST INC.
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI Racie LA, Spaulding V, Treacy M;
DR WPI; 99-070076/06.
PT New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries
PS Claim 1; Page 321; 633pp; English.
CC This sequence represents an expressed sequence tag (EST), and is a
CC polynucleotide of the invention. The polynucleotides of the invention are
CC all secreted EST sequences isolated from a variety of human tissue
CC sources. The EST sequences and proteins encoded by them are predicted to
CC have useful biological activities which would make them suitable for
CC treating, preventing or ameliorating medical conditions in humans and
CC animals, although no supporting data is given. Suggested activities
CC include nutritional activity, immune stimulating or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The EST sequences are also stated to be useful for gene
CC therapy.
SQ Sequence 448 BP; 144 A; 137 C; 52 G; 115 T;

Query Match 7.7%; Score 104.4; DB 1; Length 448;
Best Local Similarity 99.1%; Pred. No. 2.4e-18;
Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      3 gcaggcagttgaggtggattaaaccaaacccagctacgcaaaatcttagcatactcctca 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    252 GCAGGCAGTTGAGGTGGATTAAACCAAACCCAGCTACGCAAAATCTTAGCATACTCCTCA 311

Qy     63 attacccacataggatgaataatagcagttctaccgtacaaccccg 108
          |||||||||||||||||||||||||||||||||||||||||||||| |
Db    312 ATTACCCACATAGGATGAATAATAGCAGTTCTACCGTACAACCCTG 357
```

RESULT 5
V38226
ID V38226 standard; DNA; 350 BP.
AC V38226;
DT 28-SEP-1998 (first entry)
DE Human mucosal adressin cell adhesion molecule-1 gene exon 3.
KW Mucosal addressin cell adhesion molecule-1; MAdCAM-1; human;
KW cancer; tumour; inflammation; transplant rejection; arthritis;
KW rheumatoid arthritis; infection; inflammatory bowel disease;
KW autoimmune disease; experimental autoimmune encephalitis;
KW dermatosis; diagnosis; therapy; ds.
OS Homo sapiens.
PN WO9820110-A1.
PD 14-MAY-1998.
PF 01-NOV-1996; U17549.
PR 01-NOV-1996; WO-U17549.
PA (HUMA-) HUMAN GENOME SCI INC.
PA (UYAU-) UNIV AUCKLAND.
PI Greene JM, Krissansen GW, Leung EYF, Ni J, Ruben SM;
DR WPI; 98-286926/25.
PT Addressin cell adhesion molecules - used to develop products for
PT detection of inflammatory conditions or cancer and for treating or

```
PT   preventing inflammatory conditions
PS   Claim 1; Page 114; 164pp; English.
CC   This nucleic acid molecule comprises exon 3 of the novel human
CC   mucosal vascular addressin cell adhesion molecule MAdCAM-1 gene
CC   that maps to chromosome 19, band p13.3.  A MAdCAM-1 genomic clone
CC   (deposited as ATCC 97758) was isolated from a cosmid library by
CC   screening with a human EST homologue of murine MAdCAM-1.  The 5'
CC   untranslated region and exons 1-5 (see V38223-28), in various
CC   combinations, comprise the coding regions of MAdCAM-1(a) and its
CC   splice variants MAdCAM-1(b-e) (see V38201-05).  Exons 2 and 3
CC   encode the N-terminal immunoglobulin-like domains.  The invention
CC   provides MAdCAM-1(a-e) polypeptides (see W60607-11), vectors, host
CC   cells and recombinant methods of producing the polypeptides.  Also
CC   provided are diagnostic methods for detecting cancer or a
CC   pathological inflammatory condition (e.g. transplant rejection,
CC   arthritis, rheumatoid arthritis, infection, dermatosis, inflammatory
CC   bowel disease, chronic relapsing experimental autoimmune
CC   encephalitis), and therapeutic methods for treating an individual
CC   in need of a reduction in activity of any of MAdCAM-1(a-e).  The
CC   invention also relates to a method of screening compounds for the
CC   ability to regulate expression of MAdCAM-1(a-e) from its promoter.
SQ   Sequence  350 BP;  52 A;  129 C;  120 G;  49 T;

  Query Match             3.3%;  Score 45.2;  DB 1;  Length 350;
  Best Local Similarity   47.2%;  Pred. No. 0.0078;
  Matches  137;  Conservative    0;  Mismatches  153;  Indels    0;  Gaps    0;

Qy    385 ccagaggtggcagcaatggcccggctactaggggacctagacaggagcacgttcagaaag 444
          || |||||||||    | |||||    | | |  | | |||   | | || ||
Db     61 CCGGAGGTGGCCTGTACGGCCCACAAAGTCACGCCCGTGGACCCCAACGCGCTCTCCTTC 120

Qy    445 ttgctgaagtttgtggtcagcagcctgcaggggggaggactgccgagacggtgtgcagcgt 504
          |  |||    | | || |      ||| |||||| | |   ||  | |   | |  ||
Db    121 TCCCTGCTCGTCGGGGGCCAGGAACTGGAGGGGGCGCAAGCCCTGGGCCCGGAGGTGCAG 180

Qy    505 cttggggtcagcgccaacctgccggaggagcagctgggtgccctgctggcaggcatgcac 564
            | ||     |    || |  ||| ||| | ||     |  | || |||
Db    181 GAGGAGGAGGAGGAGCCCCAGGGGGACGAGGACGTGCTGTTCAGGGTGACAGAGCGCTGG 240

Qy    565 acactgctccagcaggccctccgtctgccccccaccagcctgaagcctgacaccttcagg 624
            ||||  |  | ||    || | | || ||| ||  |      |  | |||  | |||
Db    241 CGGCTGCCGCCCCTGGGGACCCCTGTCCCGCCCGCCCTCTACTGCCAGGCCACGATGAGG 300

Qy    625 gaccagctccaggagctctgcatccccccaagacctggtcggggacttggc 674
            |    |  ||||||| ||  || ||| | | |   ||| || |  ||
Db    301 CTGCCTGGCTTGGAGCTCAGCCACCGCCAGGCCATCCCCGGTGAGTCCGC 350


RESULT   6
T62138/c
ID   T62138 standard; DNA; 1771 BP.
AC   T62138;
DT   13-JUL-1997  (first entry)
DE   Leishmania tropica Lt-210 antigen cDNA.
KW   Leishmaniasis; vaccine; diagnosis; Lt-210; antigen; immunogen; ss.
OS   Leishmania tropica strain MHOM/SA/91/WR1063C.
FH   Key             Location/Qualifiers
FT   cds             1. .1701
FT                   /*tag=  a
PN   WO9711180-A1.
PD   27-MAR-1997.
PF   20-SEP-1996; U15185.
PR   22-SEP-1995; US-533669.
PA   (CORI-) CORIXA CORP.
PI   Campos-Neto A,  Dillon DC,  Reed SG,  Skeiky YA,  Webb JR;
DR   WPI; 97-202888/18.
DR   P-PSDB; W11217.
PT   Polypeptide(s) comprising immunogenic part of Leishmania antigen -
PT   useful for leishmaniasis vaccination and diagnosis
PS   Example 4; Page 64-67; 106pp; English.
CC   A genomic DNA molecule (T62138) encodes a 210 kDa immunogenic
CC   Leishmania tropica antigen (W11217), designated Lt-210.  It was
CC   obtd. by screening a L. tropica promastigote expression library
CC   with sera isolated from viscerotropic leishmaniasis patients.  DNA
CC   molecules (see also T62135-37, T62139) encoding Leishmania antigens
CC   can be incorporated into vectors and used to produce recombinant
CC   antigens, or immunogenic portions of them, in host cells for use in
CC   vaccines against Leishmaniasis, or may themselves be used as
CC   vaccines for in vivo immunogen generation.
SQ   Sequence  1771 BP;  279 A;  552 C;  672 G;  268 T;

  Query Match             3.2%;  Score 44;  DB 1;  Length 1771;
  Best Local Similarity   45.8%;  Pred. No. 0.031;
  Matches  152;  Conservative    0;  Mismatches  180;  Indels    0;  Gaps    0;

Qy    376 cagcttcctccagaggtggcagcaatggcccggctactaggggacctagacaggagcacg 435
          | ||  |||| | ||  | |  |  | || |  ||    |  | | |    | |  ||
Db   1050 CCGCGCCCTCGATAGCAGTCGCCCTTTGCACTCCTCGCTGTAGCTCGACGTCGCCGGCCG 991

Qy    436 ttcagaaagttgctgaagtttgtggtcagcagcctgcaggggggaggactgccgagacggt 495
           |||     || ||     |     ||    ||     | |  |  |||||   |  ||
Db    990 CTCATCTTTTTCCTCCTCCTCAGCCTCCAGCTCCGCGCAGAGACGCGCTGCCTCCTCGAG 931

Qy    496 gtgcagcgtcttggggtcagcgccaacctgccggaggagcagctgggtgccctgctggca 555
          | | | |    ||  |   || |  | |||  |     |||  |||   ||| || |
Db    930 GGCCTGGGAGCGGGCCTGCTCGGCCTCGTGCATGGCCTCCAGGCGGGCCGCCTCCTCGAG 871

Qy    556 ggcatgcacacactgctccagcaggccctccgtctgccccccaccagcctgaagcctgac 615
          |||  || ||||    || |      | | | |     ||   |  |||     |  | |
Db    870 GGCCTGGACACGGGCCTGCTCGCCCTCGTGCATGACATCCAGGCGGGCCGCCTCCTCGGC 811

Qy    616 accttcagggaccagctccaggagctctgcatcccccaagacctggtcggggacttggcc 675
            | ||  |  ||      ||  || |||| ||| | | | ||||| || || | ||  |
Db    810 CGCCTCCAGCTCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGACGCGGGCCTGCTC 751

Qy    676 agcgtggtatttgggagccagcggcccctcct 707
          || | ||.  ||  | ||||  |  || |||
Db    750 GGCCTCGTGCATGACATCCAGGCGGGCCGCCT 719


RESULT   7
V47558/c
ID   V47558 standard; DNA; 1771 BP.
AC   V47558;
DT   13-NOV-1998  (first entry)
DE   Leishmania antigen Lt-210 coding sequence.
KW   Leishmania antigen; immune response; infection detection; therapy;
KW   humoral response induction; cellular response induction; cancer;
KW   interleukin-12 production; ss.
OS   Leishmania sp.
FH   Key             Location/Qualifiers
FT   CDS             1. .1701
FT                   /*tag=  a
FT                   /product= Lt-210
PN   WO9835045-A2.
PD   13-AUG-1998.
PF   12-FEB-1998; U03002.
PR   27-AUG-1997; US-920609.
PR   12-FEB-1997; US-798841.
PA   (CORI-) CORIXA CORP.
PI   Campos-Neto A, Dillon DC, Reed SG, Skeiky YA, Webb JR;
DR   WPI; 98-447242/38.
DR   P-PSDB; W70212.
PT   New immunogenic fragments of Leishmania antigens and related nucleic
PT   acid, vectors and host cells - are useful for diagnosis, prevention
PT   and treatment of leishmaniasis, also to induce production of
PT   interleukin-12 generally
PS   Disclosure; Page 79-82; 19[illegible]pp; English.
CC   This sequence encodes a Leishmania antigen (LAg) of the invention,
CC   designated Lt-210. Compositions and vaccines containing the protein are
CC   used to generate a protective or therapeutic immune response against the
CC   Leishmania species donavani, chagasi, infantum, major, amazonensis,
CC   braziliensis, panamensis, tropica or guayanensis. They can also be used
CC   to detect infection (in a skin test). The compositions induce a humoral
```

```
CC   and/or cellular response, specifically of Th1 type, particularly
CC   including induction of interleukin-12 (IL-12) production. They may thus
CC   be used more generally to treat any condition (e.g. bacterial, viral or
CC   protozoal infection, or cancer) which responds to IL-12.
SQ   Sequence   1771 BP;    279 A;    552 C;    672 G;    268 T;


  Query Match             3.2%;  Score 44;  DB 1;  Length 1771;
  Best Local Similarity   45.8%;  Pred. No. 0.031;
  Matches  152;  Conservative    0;  Mismatches  180;  Indels    0;  Gaps    0;

Qy     376 cagcttcctccagaggtggcagcaatggcccggctactaggggacctagacaggagcacg 435
           | ||  |||| | ||  | |  |  | || |  ||    |  |  | |    |  |  ||
Db    1050 CCGCGCCCTCGATAGCAGTCGCCCTTTGCACTCCTCGCTGTAGCTCGACGTCGCCGGCCG 991

Qy     436 ttcagaaagttgctgaagtttgtggtcagcagcctgcaggggggaggactgccgagacggt 495
           |||     || ||     |     ||     ||     | |  |  |||||    ||
Db     990 CTCATCTTTTTCCTCCTCCTCAGCCTCCAGCTCCGCGCAGAGACGCGCTGCCTCCTCGAG 931

Qy     496 gtgcagcgtcttggggtcagcgccaacctgccggaggagcagctgggtgccctgctggca 555
           | | | |    ||  |   || |  | |||  |     |||  |||   |||  || |
Db     930 GGCCTGGGAGCGGGCCTGCTCGGCCTCGTGCATGGCCTCCAGGCGGGCCGCCTCCTCGAG 871

Qy     556 ggcatgcacacactgctccagcaggccctccgtctgccccccaccagcctgaagcctgac 615
           |||  || ||||    || |      | | | |     ||   |  |||     |  | |
Db     870 GGCCTGGACACGGGCCTGCTCGCCCTCGTGCATGACATCCAGGCGGGCCGCCTCCTCGGC 811

Qy     616 accttcagggaccagctccaggagctctgcatcccccaagacctggtcggggacttggcc 675
             | ||  |  ||      ||  || |||| ||| | | | |||||  || || | ||  |
Db     810 CGCCTCCAGCTCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGACGCGGGCCTGCTC 751

Qy     676 agcgtggtatttgggagccagcggcccctcct 707
           || | ||   ||  | ||||  |  || |||
Db     750 GGCCTCGTGCATGACATCCAGGCGGGCCGCCT 719


RESULT   8
V38201
ID   V38201 standard; DNA; 1536 BP.
AC   V38201;
DT   28-SEP-1998  (first entry)
DE   Human mucosal adressin cell adhesion molecule-1(a) DNA.
KW   Mucosal addressin cell adhesion molecule-1; MAdCAM-1(a); human;
KW   cancer; tumour; inflammation; transplant rejection; arthritis;
KW   rheumatoid arthritis; infection; inflammatory bowel disease;
KW   autoimmune disease; experimental autoimmune encephalitis;
KW   dermatosis; diagnosis; therapy; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             1. .1149
FT                   /*tag=  a
FT   sig_peptide     1. .49
FT                   /*tag=  b
FT   mat_peptide     52. .1146
FT                   /*tag=  c
FT   polyA_signal    1507. .1512
FT                   /*tag=  d
PN   WO9820110-A1.
PD   14-MAY-1998.
PF   01-NOV-1996; U17549.
PR   01-NOV-1996; WO-U17549.
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (UYAU-) UNIV AUCKLAND.
PI   Greene JM, Krissansen GW, Leung EYF, Ni J, Ruben SM;
DR   WPI; 98-286926/25.
DR   P-PSDB; W60607.
PT   Addressin cell adhesion molecules  - used to develop products for
PT   detection of inflammatory conditions or cancer and for treating or
PT   preventing inflammatory conditions
PS   Claim 2; Page 88-90; 164pp; English.
CC   This nucleic acid molecule, designated clone HEBBC23 and deposited
CC   as ATCC 97759, codes for human mucosal vascular addressin cell
CC   adhesion molecule MAdCAM-1(a) (see W60607), a novel cell surface
CC   adhesion molecule that shows homology to murine MAdCAM-1.  The
CC   invention relates to human MAdCAM-1(a) as well as 4 splice variants,
CC   designated MAdCAM-1(b), -1(c), -1(d) and -1(e) (see V38201-05).
CC   These 5 nucleic acid molecules were discovered in a cDNA library
CC   derived from human foetal brain cells following a database search
CC   for human ESTs having homology for mouse MAdCAM-1.  The genes were
CC   also identified in cDNA libraries from the small intestine, colon,
CC   spleen and pancreas.  The invention also provides MAdCAM-1(a-e)
CC   proteins (see W60607-11), vectors, host cells, recombinant methods
CC   of producing the polypeptides, as well as methods for identifying
CC   agonists and antagonists of activity, diagnostic methods for
CC   detecting cancer or a pathological inflammatory condition, and
CC   therapeutic methods for treating an individual in need of a
CC   reduction in activity of any of MAdCAM-1(a-e).  Genomic DNA
CC   molecules comprising the 5' untranslated region and exons 1-5 (see
CC   V38223-28) which, in various combinations, comprise the coding
CC   region of any of the MAdCAM-1 splice variants are also claimed.
CC   The novel human MAdCAM-1 polypeptides can be used as a target for
CC   the diagnosis and treatment of inflammation conditions such as
CC   transplant rejection, arthritis, rheumatoid arthritis, infection,
CC   dermatosis, inflammatory bowel disease, and autoimmune disease,
CC   including chronic relapsing experimental autoimmune encephalitis.
SQ   Sequence   1536 BP;    263 A;    571 C;    442 G;    260 T;


  Query Match             3.1%;  Score 41.8;  DB 1;  Length 1536;
  Best Local Similarity   46.4%;  Pred. No. 0.11;
  Matches  136;  Conservative    0;  Mismatches  157;  Indels    0;  Gaps    0;

Qy     372 ggcccagcttcctccagaggtggcagcaatggcccggctactaggggacctagacaggag 431
           || ||    |   || ||||||||    | |||||    | |   |  | | |||    |
Db     375 GGTCCCTGGTGACCCGGAGGTGGCCTGTACGGCCCACAAAGTCACGCCCGTGGACCCCAA 434

Qy     432 cacgttcagaaagttgctgaagtttgtggtcagcagcctgcaggggggaggactgccgaga 491
           | || ||      | |||    | | || |      ||| |||||| | |   ||  |
Db     435 CGCGCTCTCCTTCTCCCTGCTCGTCGGGGGCCAGGAACTGGAGGGGGCGCAAGCCCTGGG 494

Qy     492 cggtgtgcagcgtcttggggtcagcgccaacctgccggaggagcagctgggtgccctgct 551
           |   | | ||     | ||     |    || |  ||| ||| | |  ||     |  | |
Db     495 CCCGGAGGTGCAGGAGGAGGAGGAGGAGCCCCAGGGGGACGAGGACGTGCTGTTCAGGGT 554

Qy     552 ggcaggcatgcacacactgctccagcaggccctccgtctgccccccaccagcctgaagcc 611
           | |||           ||||  | ||    || | | || ||| ||  |      |
Db     555 GACAGAGCGCTGGCGGCTGCCGCCCCTGGGGACCCCTGTCCCGCCCGCCCTCTACTGCCA 614

Qy     612 tgacaccttcagggaccagctccaggagctctgcatcccccaagacctggtcg 664
           | |||  | |||   |    |  |||||| ||  || ||| | | |   ||
Db     615 GGCCACGATGAGGCTGCCTGGCTTGGAGCTCAGCCACCGCCAGGCCATCCCCG 667


RESULT   9
V38202
ID   V38202 standard; DNA; 1488 BP.
AC   V38202;
DT   28-SEP-1998  (first entry)
DE   Human mucosal adressin cell adhesion molecule-1(b) DNA.
KW   Mucosal addressin cell adhesion molecule-1; MAdCAM-1(b); human;
KW   cancer; tumour; inflammation; transplant rejection; arthritis;
KW   rheumatoid arthritis; infection; inflammatory bowel disease;
KW   autoimmune disease; experimental autoimmune encephalitis;
KW   dermatosis; diagnosis; therapy; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             1. .1101
FT                   /*tag=  a
FT   sig_peptide     1. .49
FT                   /*tag=  b
FT   mat_peptide     52. .1098
FT                   /*tag=  c
FT   polyA_signal    1459. .1464
FT                   /*tag=  d
PN   WO9820110-A1.
PD   14-MAY-1998.
```

```
PF   01-NOV-1996; U17549.
PR   01-NOV-1996; WO-U17549.
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (UYAU-) UNIV AUCKLAND.
PI   Greene JM, Krissansen GW, Leung EYF, Ni J, Ruben SM;
DR   WPI; 98-286926/25.
DR   P-PSDB; W60608.
PT   Addressin cell adhesion molecules  - used to develop products for
PT   detection of inflammatory conditions or cancer and for treating or
PT   preventing inflammatory conditions
PS   Claim 2; Page 92-94; 164pp; English.
CC   This nucleic acid molecule, designated clone HSKCW36, codes for
CC   human mucosal vascular addressin cell adhesion molecule MAdCAM-1(b)
CC   (see W60608), a novel cell surface adhesion molecule that shows
CC   homology to murine MAdCAM-1.  The invention relates to human
CC   MAdCAM-1(a) as well as 4 splice variants, designated MAdCAM-1(b),
CC   -1(c), -1(d) and -1(e) (see V38201-05).  These 5 nucleic acid
CC   molecules were discovered in a cDNA library derived from human
CC   foetal brain cells following a database search for human ESTs
CC   having homology for mouse MAdCAM-1.  The genes were also identified
CC   in cDNA libraries from the small intestine, colon, spleen and
CC   pancreas.  The invention also provides MAdCAM-1(a-e) proteins (see
CC   W60607-11), vectors, host cells, recombinant methods of producing
CC   the polypeptides, as well as methods for identifying agonists and
CC   antagonists of activity, diagnostic methods for detecting cancer or
CC   a pathological inflammatory condition, and therapeutic methods for
CC   treating an individual in need of a reduction in activity of any of
CC   MAdCAM-1(a-e).  Genomic DNA molecules comprising the 5' untranslated
CC   region and exons 1-5 (see V38223-28) which, in various combinations,
CC   comprise the coding region of any of the MAdCAM-1 splice variants
CC   are also claimed.  The novel human MAdCAM-1 polypeptides can be
CC   used as a target for the diagnosis and treatment of inflammation
CC   conditions such as transplant rejection, arthritis, rheumatoid
CC   arthritis, infection, dermatosis, inflammatory bowel disease, and
CC   autoimmune disease, including chronic relapsing experimental
CC   autoimmune encephalitis.
SQ   Sequence   1488 BP;    255 A;    543 C;    435 G;    255 T;

  Query Match             3.1%;  Score 41.6;  DB 1;  Length 1488;
  Best Local Similarity   46.8%;  Pred. No. 0.12;
  Matches  131;  Conservative    0;  Mismatches  149;  Indels    0;  Gaps    0;

Qy    385 ccagaggtggcagcaatggcccggctactaggggacctagacaggagcacgttcagaaag 444
          || ||||||||    | |||||    | |   |  | | |||   | | || ||
Db    388 CCGGAGGTGGCCTGTACGGCCCACAAAGTCACGCCCGTGGACCCCAACGCGCTCTCCTTC 447

Qy    445 ttgctgaagtttgtggtcagcagcctgcaggggggaggactgccgagacggtgtgcagcgt 504
          |  |||    | | || |      ||| |||||| | |   ||  | |   | |  ||
Db    448 TCCCTGCTCGTCGGGGGCCAGGAACTGGAGGGGGCGCAAGCCCTGGGCCCGGAGGTGCAG 507

Qy    505 cttggggtcagcgccaacctgccggaggagcagctgggtgccctgctggcaggcatgcac 564
            | ||      |    || |  ||| ||| |  ||      |  | || |||
Db    508 GAGGAGGAGGAGGAGCCCCAGGGGGACGAGGACGTGCTGTTCAGGGTGACAGAGCGCTGG 567

Qy    565 acactgctccagcaggccctccgtctgccccccaccagcctgaagcctgacaccttcagg 624
            ||||  |  | ||    || | | || ||| ||  |      |  | |||  | |||
Db    568 CGGCTGCCGCCCCTGGGGACCCCTGTCCCGCCCGCCCTCTACTGCCAGGCCACGATGAGG 627

Qy    625 gaccagctccaggagctctgcatcccccaagacctggtcg 664
            |    |  ||||||| ||  || ||| | | |   ||
Db    628 CTGCCTGGCTTGGAGCTCAGCCACCGCCAGGCCATCCCCG 667


RESULT  10
T39751
ID   T39751 standard; cDNA; 1539 BP.
AC   T39751;
DT   01-JAN-1997  (first entry)
DE   Human mucosal addressin cell adhesion molecule-1 cDNA clone 20.
KW   Mucosal addressin cell adhesion molecule-1; MAdCAM; antibody;
KW   leukocyte infiltration; ulcerative colitis; Crohn's disease; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             1. .1149
FT                   /*tag= a
FT   signal_peptide  1. .54
FT                   /*tag= b
FT   mat_peptide     55. .1146
FT                   /*tag= c
PN   WO9624673-A1.
PD   15-AUG-1996.
PF   12-FEB-1996; U02153.
PR   10-FEB-1995; US-386857.
PR   01-SEP-1995; US-523004.
PA   (LEUK-) LEUKOSITE INC.
PI   Briskin MJ,  Newman W,  Picarella D,  Ringler DJ;
DR   WPI; 96-384445/38.
DR   P-PSDB; W05322.
PT   Nucleic acid encoding primate mucosal addressin cell adhesion
PT   molecule-1 - and antibodies against this protein, useful to treat
PT   diseases associated with leukocyte infiltration e.g. ulcerative
PT   colitis
PS   Claim 3; Page 108-110; 156pp; English.
CC   cDNA clone 20 (T39751) codes for human mucosal addressin cell
CC   adhesion molecule-1 (MAdCAM-1) (W05322), whose cellular adhesion
CC   function is alpha4beta7 integrin-dependent and pref. is selective
CC   for alpha4beta7.  It was isolated from a cDNA library prepd. from
CC   mesenteric lymph nodes (MLN) of a patient with Crohn's disease,
CC   using a macaque MAdCAM-1 cDNA clone (T39752) as probe.  Another
CC   clone (T39750), encoding a different human MAdCAM isoform (W05321),
CC   was isolated from a histologically normal MLN library.  MAdCAM
CC   nucleic acids can be used as probes to detect polymorphic or
CC   allelic variants of MAdCAM, or used to produce recombinant MAdCAMs
CC   in host cells.
SQ   Sequence   1539 BP;    268 A;    570 C;    442 G;    259 T;

  Query Match             3.1%;  Score 41.6;  DB 1;  Length 1539;
  Best Local Similarity   46.8%;  Pred. No. 0.13;
  Matches  131;  Conservative    0;  Mismatches  149;  Indels    0;  Gaps    0;

Qy    385 ccagaggtggcagcaatggcccggctactaggggacctagacaggagcacgttcagaaag 444
          || ||||||||    | |||||    | |   |  | | |||   | | || ||
Db    388 CCGGAGGTGGCCTGTACGGCCCACAAAGTCACGCCCGTGGACCCCAACGCGCTCTCCTTC 447

Qy    445 ttgctgaagtttgtggtcagcagcctgcaggggggaggactgccgagacggtgtgcagcgt 504
          |  |||    | | || |      ||| |||||| | |   ||  | |   | |  ||
Db    448 TCCCTGCTCGTCGGGGGCCAGGAACTGGAGGGGGCGCAAGCCCTGGGCCCGGAGGTGCAG 507

Qy    505 cttggggtcagcgccaacctgccggaggagcagctgggtgccctgctggcaggcatgcac 564
            | ||      |    || |  ||| ||| |  ||      |  | || |||
Db    508 GAGGAGGAGGAGGAGCCCCAGGGGGACGAGGACGTGCTGTTCAGGGTGACAGAGCGCTGG 567

Qy    565 acactgctccagcaggccctccgtctgccccccaccagcctgaagcctgacaccttcagg 624
            ||||  |  | ||    || | | || ||| ||  |      |  | |||  | |||
Db    568 CGGCTGCCGCCCCTGGGGACCCCTGTCCCGCCCGCCCTCTACTGCCAGGCCACGATGAGG 627

Qy    625 gaccagctccaggagctctgcatcccccaagacctggtcg 664
            |    |  ||||||| ||  || ||| | | |   ||
Db    628 CTGCCTGGCTTGGAGCTCAGCCACCGCCAGGCCATCCCCG 667


RESULT  11
X35280
ID   X35280 standard; DNA; 1539 BP.
AC   X35280;
DT   02-JUL-1999  (first entry)
DE   cDNA encoding human MAdCAM-1.
KW   Mucosal addressing cell adhesion molecule-1; MAdCAM-1; inhibitor;
KW   leukocyte infiltration; disease; inflammatory bowel disease;
KW   insulin-dependent diabetes mellitus; inflammatory bowel disease;
KW   ulcerative colitis; Crohn's disease; ileitis' coeliac disease;
KW   nontropical Sprue; enteropathy; seronegative arthropathy; colitis;
KW   eosinophilic gastroenteritis; pouchitis; proctocolectomy;
KW   ileoanal anastomosis; alpha4-beta-integrin; ss.
```

```
OS   Homo sapiens.
PN   WO9725351-A2.
PD   17-JUL-1997.
PF   03-JAN-1997; U00291.
PR   04-JAN-1996; US-582740.
PA   (LEUK-) LEUKOSITE INC.
PI   Schwender CF, Shroff HN;
DR   WPI; 97-384978/35.
DR   P-PSDB; Y02069.
PT   New polypeptide(s) for inhibiting MAdCAM-1-mediated interactions -
PT   useful for treatment of diseases associated with leukocyte
PT   infiltration of tissues, especially inflammatory bowel disease and
PT   insulin-dependent diabetes mellitus
PS   Disclosure; Fig 5; 108pp; English.
CC   The specification describes inhibitors of mucosal addressing cell
CC   adhesion molecule-1 (MAdCAM-1) mediated interactions. These inhibitors
CC   comprise peptides which mimic the conserved amino acid motif LDTSL of
CC   MAdCAM-1, and which have groups bonded to the N and C terminals.
CC   The inhibitors can be used for the treatment of a disease associated
CC   with leukocyte infiltration of tissues expressing the MAdCAM-1,
CC   especially inflammatory bowel disease and insulin-dependent diabetes
CC   mellitus. Inflammatory bowel diseases which can be treated include
CC   ulcerative colitis, Crohn's disease, ileitis, coeliac disease,
CC   nontropical Sprue, enteropathy associated with seronegative
CC   arthropathies, microscopic or collagenous colitis, eosinophilic
CC   gastroenteritis, or pouchitis resulting after proctocolectomy, and
CC   ileoanal anastomosis. The inhibitors can also be used to inhibit the
CC   bonding of a cell expressing a ligand for MAdCAM-1 on the cell surface to
CC   MAdCAM-1. They are also useful in diagnostic and research applications,
CC   e.g. as immunogens (when conjugated to a suitable carrier) to induce the
CC   formation of antibodies which selectively bind MAdCAM-1. These antibodies
CC   can in turn be used to identify cells expressing MAdCAM-1 on their cell
CC   surface or to detect MAdCAM-1 in a sample. The compounds can also be
CC   labelled and used to detect alpha4-beta-integrin and/or to quantitate
CC   expression of this the surface of cells. The present sequence
CC   encodes human MAdCAM-1.
SQ   Sequence   1539 BP;    268 A;    570 C;    442 G;    259 T;

  Query Match             3.1%;  Score 41.6;  DB 1;  Length 1539;
  Best Local Similarity   46.8%;  Pred. No. 0.13;
  Matches  131;  Conservative    0;  Mismatches  149;  Indels    0;  Gaps    0;

Qy    385 ccagaggtggcagcaatggcccggctactaggggacctagacaggagcacgttcagaaag 444
          || |||||||||    | |||||    | |   |  | | |||   | | || ||
Db    388 CCGGAGGTGGCCTGTACGGCCCACAAAGTCACGCCCGTGGACCCCAACGCGCTCTCCTTC 447

Qy    445 ttgctgaagtttgtggtcagcagcctgcaggggggaggactgccgagacggtgtgcagcgt 504
          |  |||    | | || |     ||| |||||| | |   ||  | |   | |  ||
Db    448 TCCCTGCTCGTCGGGGGCCAGGAACTGGAGGGGGCGCAAGCCCTGGGCCCGGAGGTGCAG 507

Qy    505 cttggggtcagcgccaacctgccggaggagcagctgggtgccctgctggcaggcatgcac 564
            | ||      |    || |  ||| ||| |  ||      |  | || |||
Db    508 GAGGAGGAGGAGGAGCCCCAGGGGGACGAGGACGTGCTGTTCAGGGTGACAGAGCGCTGG 567

Qy    565 acactgctccagcaggccctccgtctgccccccaccagcctgaagcctgacaccttcagg 624
          ||||  |  | ||    || | | || ||| ||  |       |  | |||  | |||
Db    568 CGGCTGCCGCCCCTGGGGACCCCTGTCCCGCCCGCCCTCTACTGCCAGGCCACGATGAGG 627

Qy    625 gaccagctccaggagctctgcatcccccaagacctggtcg 664
            |    |  |||||||| ||  || ||| | | |   ||
Db    628 CTGCCTGGCTTGGAGCTCAGCCACCGCCAGGCCATCCCCG 667


RESULT  12
T39750
ID   T39750 standard; cDNA; 1624 BP.
AC   T39750;
DT   01-JAN-1997  (first entry)
DE   Human mucosal addressin cell adhesion molecule-1 cDNA clone 4.
KW   Mucosal addressin cell adhesion molecule-1; MAdCAM; antibody;
KW   leukocyte infiltration; ulcerative colitis; Crohn's disease; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds.            1. .1221
FT                   /*tag=  a
FT   signal_peptide  1. .54
FT                   /*tag=  b
FT   mat_peptide     55. .1218
FT                   /*tag=  c
PN   WO9624673-A1.
PD   15-AUG-1996.
PF   12-FEB-1996; U02153.
PR   10-FEB-1995; US-386857.
PR   01-SEP-1995; US-523004.
PA   (LEUK-) LEUKOSITE INC.
PI   Briskin MJ,  Newman W,  Picarella D,  Ringler DJ;
DR   WPI; 96-384445/38.
DR   P-PSDB; W05321.
PT   Nucleic acid encoding primate mucosal addressin cell adhesion
PT   molecule-1 - and antibodies against this protein, useful to treat
PT   diseases associated with leukocyte infiltration e.g. ulcerative
PT   colitis
PS   Claim 3; Page 104-106; 156pp; English.
CC   cDNA clone 4 (T39750) codes for human mucosal addressin cell
CC   adhesion molecule-1 (MAdCAM-1) (W05321), whose cellular adhesion
CC   function is alpha4beta7 integrin-dependent and pref. is selective
CC   for alpha4beta7.  It was isolated from a cDNA library prepd. from
CC   normal human mesenteric lymph nodes (MLN) using a macaque MAdCAM-1
CC   cDNA clone (T39752) as probe.  Another clone (T39751), encoding
CC   a different human MAdCAM isoform (W05322), was isolated from the
CC   MLN of a patient with Crohn's disease.  MAdCAM nucleic acids can
CC   be used as probes to detect polymorphic or allelic variants of
CC   MAdCAM, or used to produce recombinant MAdCAMs in host cells.
SQ   Sequence   1624 BP;    295 A;    610 C;    454 G;    265 T;

  Query Match             3.1%;  Score 41.6;  DB 1;  Length 1624;
  Best Local Similarity   46.8%;  Pred. No. 0.13;
  Matches  131;  Conservative    0;  Mismatches  149;  Indels    0;  Gaps    0;

Qy    385 ccagaggtggcagcaatggcccggctactaggggacctagacaggagcacgttcagaaag 444
          || |||||||||    | |||||    | |   |  | | |||   | | || ||
Db    388 CCGGAGGTGGCCTGTACGGCCCACAAAGTCACGCCCGTGGACCCCAACGCGCTCTCCTTC 447

Qy    445 ttgctgaagtttgtggtcagcagcctgcaggggggaggactgccgagacggtgtgcagcgt 504
          |  |||    | | || |     ||| |||||| | |   ||  | |   | |  ||
Db    448 TCCCTGCTCGTCGGGGGCCAGGAACTGGAGGGGGCGCAAGCCCTGGGCCCGGAGGTGCAG 507

Qy    505 cttggggtcagcgccaacctgccggaggagcagctgggtgccctgctggcaggcatgcac 564
            | ||      |    || |  ||| ||| |  ||      |  | || |||
Db    508 GAGGAGGAGGAGGAGCCCCAGGGGGACGAGGACGTGCTGTTCAGGGTGACAGAGCGCTGG 567

Qy    565 acactgctccagcaggccctccgtctgccccccaccagcctgaagcctgacaccttcagg 624
          ||||  |  | ||    || | | || ||| ||  |       |  | |||  | |||
Db    568 CGGCTGCCGCCCCTGGGGACCCCTGTCCCGCCCGCCCTCTACTGCCAGGCCACGATGAGG 627

Qy    625 gaccagctccaggagctctgcatcccccaagacctggtcg 664
            |    |  |||||||| ||  || ||| | | |   ||
Db    628 CTGCCTGGCTTGGAGCTCAGCCACCGCCAGGCCATCCCCG 667


RESULT  13
X35279
ID   X35279 standard; DNA; 1624 BP.
AC   X35279;
DT   02-JUL-1999  (first entry)
DE   cDNA encoding human MAdCAM-1.
KW   Mucosal addressing cell adhesion molecule-1; MAdCAM-1; inhibitor;
KW   leukocyte infiltration; disease; inflammatory bowel disease;
KW   insulin-dependent diabetes mellitus; inflammatory bowel disease;
KW   ulcerative colitis; Crohn's disease; ileitis' coeliac disease;
KW   nontropical Sprue; enteropathy; seronegative arthropathy; colitis;
KW   eosinophilic gastroenteritis; pouchitis; proctocolectomy;
KW   ileoanal anastomosis; alpha4-beta-integrin; ss.
OS   Homo sapiens.
```

PN WO9725351-A2.
PD 17-JUL-1997.
PF 03-JAN-1997; U00291.
PR 04-JAN-1996; US-582740.
PA (LEUK-) LEUKOSITE INC.
PI Schwender CF, Shroff HN;
DR WPI; 97-384978/35.
DR P-PSDB; Y02068.
PT New polypeptide(s) for inhibiting MAdCAM-1-mediated interactions -
PT useful for treatment of diseases associated with leukocyte
PT infiltration of tissues, especially inflammatory bowel disease and
PT insulin-dependent diabetes mellitus
PS Disclosure; Fig 4; 108pp; English.
CC The specification describes inhibitors of mucosal addressing cell
CC adhesion molecule-1 (MAdCAM-1) mediated interactions. These inhibitors
CC comprise peptides which mimic the conserved amino acid motif LDTSL of
CC MAdCAM-1, and which have groups bonded to the N and C terminals.
CC The inhibitors can be used for the treatment of a disease associated
CC with leukocyte infiltration of tissues expressing the MAdCAM-1,
CC especially inflammatory bowel disease and insulin-dependent diabetes
CC mellitus. Inflammatory bowel diseases which can be treated include
CC ulcerative colitis, Crohn's disease, ileitis, coeliac disease,
CC nontropical Sprue, enteropathy associated with seronegative
CC arthropathies, microscopic or collagenous colitis, eosinophilic
CC gastroenteritis, or pouchitis resulting after proctocolectomy, and
CC ileoanal anastomosis. The inhibitors can also be used to inhibit the
CC bonding of a cell expressing a ligand for MAdCAM-1 on the cell surface to
CC MAdCAM-1. They are also useful in diagnostic and research applications,
CC e.g. as immunogens (when conjugated to a suitable carrier) to induce the
CC formation of antibodies which selectively bind MAdCAM-1. These antibodies
CC can in turn be used to identify cells expressing MAdCAM-1 on their cell
CC surface or to detect MAdCAM-1 in a sample. The compounds can also be
CC labelled and used to detect alpha4-beta-integrin and/or to quantitate
CC expression of this the surface of cells. The present sequence
CC encodes human MAdCAM-1.
SQ Sequence 1624 BP; 295 A; 610 C; 454 G; 265 T;

Query Match 3.1%; Score 41.6; DB 1; Length 1624;
Best Local Similarity 46.8%; Pred. No. 0.13;
Matches 131; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

```
Qy    385 ccagaggtggcagcaatggcccggctactaggggacctagacaggagcacgttcagaaag 444
          || ||||||||    | |||||    | |   |  | | |||   | | || ||
Db    388 CCGGAGGTGGCCTGTACGGCCCACAAAGTCACGCCCGTGGACCCCAACGCGCTCTCCTTC 447

Qy    445 ttgctgaagtttgtggtcagcagcctgcagggggaggactgccgagacggtgtgcagcgt 504
          |  |||    | | || |     ||| ||||||  | |   ||  | |   | |  ||
Db    448 TCCCTGCTCGTCGGGGGCCAGGAACTGGAGGGGGCGCAAGCCCTGGGCCCGGAGGTGCAG 507

Qy    505 cttggggtcagcgccaacctgccggaggagcagctgggtgccctgctggcaggcatgcac 564
            | ||      |    || |  ||| ||| |  ||     |  | || |||
Db    508 GAGGAGGAGGAGGAGCCCCAGGGGGACGAGGACGTGCTGTTCAGGGTGACAGAGCGCTGG 567

Qy    565 acactgctccagcaggccctccgtctgccccccaccagcctgaagcctgacaccttcagg 624
            ||||  |  | ||    || | | || ||| ||  |      |  | |||  | |||
Db    568 CGGCTGCCGCCCCTGGGGACCCCTGTCCCGCCCGCCCTCTACTGCCAGGCCACGATGAGG 627

Qy    625 gaccagctccaggagctctgcatcccccaagacctggtcg 664
            |    |  ||||||| ||  || ||| | | |   ||
Db    628 CTGCCTGGCTTGGAGCTCAGCCACCGCCAGGCCATCCCCG 667
```

RESULT 14
X17775
ID X17775 standard; cDNA; 1338 BP.
AC X17775;
DT 12-MAY-1999 (first entry)
DE HTLAR33 encoding cDNA (EST derived sequence).
KW YAK-1; serine-threonine protein kinase; HTLAR33; bone loss; ARDS;
KW inflammatory disorder; osteoporosis; Adult Respiratory Disease Syndrome;
KW arthritis; psoriasis; dermatitis; asthma; allergy; infection; HIV-1;
KW HIV-2; cachexia; immunodeficient disorder; septic shock; pain; injury;
KW cancer; anorexia; bulimia; Parkinson's disease; cardiovascular disease;
KW neurological disorder; Huntington's disease; gene therapy; gene mapping;
KW Gilles de la Tourette's syndrome; ss.
OS Homo sapiens.
PN EP-894863-A1.
PD 03-FEB-1999.
PF 21-JUL-1998; 305794.
PR 20-FEB-1998; US-027064.
PR 28-JUL-1997; US-053924.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PI Bergsma DJ, Shabon U;
DR WPI; 99-108353/10.
DR P-PSDB; W94998.
PT New serine-threonine kinase (HTLAR33) polypeptides and
PT polynucleotides - useful as diagnostic reagents and for prevention
PT and treatment of bone loss, neurological and inflammatory disorders
PT and cancer, HIV infections and angina pectoris
PS Claim 13; Page 15-17; 30pp; English.
CC The invention relates to a YAK-1 related serine-threonine protein kinase
CC polypeptide (HTLAR33). Host cells transformed with an expression vector
CC comprising the HTLAR33 nucleic acid are used for the recombinant
CC production of the protein. HTLAR33 polynucleotides and polypeptides are
CC useful for are useful for diagnosing susceptibility to diseases and for
CC screening for antagonists, agonists. These can be used in treatment to
CC enhance (agonist) or block (antagonist) HTLAR33 activity. Diseases
CC diagnosed, prevented or treated include: bone loss and inflammatory
CC disorders including osteoporosis, Adult Respiratory Disease Syndrome
CC (ARDS), Rheumatoid arthritis, osteoarthritis, Inflammatory Bowel Disease
CC (IBD), psoriasis, dermatitis, asthma, allergies; bacterial, fungal,
CC protozoan and viral infections, especially those caused by HIV-1 or
CC HIV-2; HIV-associated cachexia and other immunodeficient disorders;
CC septic shock; pain; injury; cancers; anorexia; bulimia; Parkinson's
CC disease; cardiovascular disease including restenosis, atherosclerosis,
CC acute heart failure, myocardial infarction; hypotension; hypertension;
CC urinary retention; angina pectoris; ulcers; benign prostatic hypertrophy;
CC and psychotic and neurological disorders, including schizophrenia, manic
CC depression, anxiety; delirium, dementia, severe mental retardation and
CC dyskinesias, such as Huntington's disease or Gilles de la Tourette's
CC syndrome. HTLAR33 polypeptides are also useful for mapping genes to
CC chromosomes, allowing gene inheritance to be studied through linkage
CC analysis.
SQ Sequence 1338 BP; 238 A; 495 C; 373 G; 232 T;

Query Match 3.0%; Score 40.2; DB 1; Length 1338;
Best Local Similarity 53.5%; Pred. No. 0.28;
Matches 84; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

```
Qy    526 ccggaggagcagctgggtgccctgctggcaggcatgcacacactgctccagcaggccctc 585
          ||||||  |  |   |    |||| |||  ||| |||  ||    | ||| |   |||
Db    719 CCGGAGCTGATGGATGTGAGCCTGGTGGGCGGCCTGCTGACTGCTCCCCACCTCACCCAG 778

Qy    586 cgtctgccccccaccagcctgaagcctgacaccttcagggaccagctccaggagctctgc 645
          || |||||||||| || || |  |||| |  ||| | |   |       | |||| ||  |
Db    779 CGCCTGCCCCCCAGCACCCGGCTGCCTCAGCCCTCCGGACTCGGATGACTGGAGGTCGTC 838

Qy    646 atccccccaagacctggtcggggacttggccagcgtgg 682
            ||||     |||  |   |  | ||  |   ||
Db    839 CACCCCTCCCGCCTCCTGATGACCCTGCCACTCTGGG 875
```

RESULT 15
X17774
ID X17774 standard; cDNA; 2394 BP.
AC X17774;
DT 12-MAY-1999 (first entry)
DE YAK-1 related serine/threonine protein kinase-HTLAR33 encoding cDNA.
KW YAK-1; serine-threonine protein kinase; HTLAR33; bone loss; ARDS;
KW inflammatory disorder; osteoporosis; Adult Respiratory Disease Syndrome;
KW arthritis; psoriasis; dermatitis; asthma; allergy; infection; HIV-1;
KW HIV-2; cachexia; immunodeficient disorder; septic shock; pain; injury;
KW cancer; anorexia; bulimia; Parkinson's disease; cardiovascular disease;
KW neurological disorder; Huntington's disease; gene therapy; gene mapping;

KW Gilles de la Tourette's syndrome; ss.
OS Homo sapiens.
PN EP-894863-A1.
PD 03-FEB-1999.
PF 21-JUL-1998; 305794.
PR 20-FEB-1998; US-027064.
PR 28-JUL-1997; US-053924.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PI Bergsma DJ, Shabon U;
DR WPI; 99-108353/10.
DR P-PSDB; W94997.
PT New serine-threonine kinase (HTLAR33) polypeptides and
PT polynucleotides - useful as diagnostic reagents and for prevention
PT and treatment of bone loss, neurological and inflammatory disorders
PT and cancer, HIV infections and angina pectoris
PS Claim 2; Page 13-15; 30pp; English.
CC This cDNA encodes a YAK-1 related serine-threonine protein kinase
CC polypeptide (HTLAR33). Host cells transformed with an expression vector
CC comprising the HTLAR33 nucleic acid are used for the recombinant
CC production of the protein. HTLAR33 polynucleotides and polypeptides are
CC useful for are useful for diagnosing susceptibility to diseases and for
CC screening for antagonists, agonists. These can be used in treatment to
CC enhance (agonist) or block (antagonist) HTLAR33 activity. Diseases
CC diagnosed, prevented or treated include: bone loss and inflammatory
CC disorders including osteoporosis, Adult Respiratory Disease Syndrome
CC (ARDS), Rheumatoid arthritis, osteoarthritis, Inflammatory Bowel Disease
CC (IBD), psoriasis, dermatitis, asthma, allergies; bacterial, fungal,
CC protozoan and viral infections, especially those caused by HIV-1 or
CC HIV-2; HIV-associated cachexia and other immunodeficient disorders;
CC septic shock; pain; injury; cancers; anorexia; bulimia; Parkinson's
CC disease; cardiovascular disease including restenosis, atherosclerosis,
CC acute heart failure, myocardial infarction; hypotension; hypertension;
CC urinary retention; angina pectoris; ulcers; benign prostatic hypertrophy;
CC and psychotic and neurological disorders, including schizophrenia, manic
CC depression, anxiety; delirium, dementia, severe mental retardation and
CC dyskinesias, such as Huntington's disease or Gilles de la Tourette's
CC syndrome. HTLAR33 polypeptides are also useful for mapping genes to
CC chromosomes, allowing gene inheritance to be studied through linkage
CC analysis.
SQ Sequence 2394 BP; 478 A; 801 C; 676 G; 439 T;

Query Match 3.0%; Score 40.2; DB 1; Length 2394;
Best Local Similarity 53.5%; Pred. No. 0.35;
Matches 84; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

```
Qy    526 ccggaggagcagctgggtgccctgctggcaggcatgcacacactgctccagcaggccctc 585
          ||||||  |  |   |    |||| |||  ||| |||  ||    | ||| |   |||
Db   1775 CCGGAGCTGATGGATGTGAGCCTGGTGGGCGGCCTGCTGACTGCTCCCCACCTCACCCAG 1834

Qy    586 cgtctgccccccaccagcctgaagcctgacaccttcagggaccagctccaggagctctgc 645
          || ||||||||||  || || |  |||| |  ||| | |   |      | |||| ||  |
Db   1835 CGCCTGCCCCCCAGCACCCGGCTGCCTCAGCCCTCCGGACTCGGATGACTGGAGGTCGTC 1894

Qy    646 atcccccaagacctggtcggggacttggccagcgtgg 682
            ||||     |||  |   |  | || |   |  ||
Db   1895 CACCCCTCCCGCCTCCTGATGACCCTGCCACTCTGGG 1931
```

Search completed: May 15, 2000, 01:46:13
Job time: 9837 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: May 15, 2000, 01:50:08 ; Search time 76.12 Seconds
(without alignments)
2260.217 Million cell updates/sec

Title: US-09-223-796-3
Perfect score: 1355
Sequence: 1 gggcaggcagttgaggtgga.....gtgtttcaggcagggcccgg 1355

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 226296 seqs, 63486255 residues

Total number of hits satisfying chosen parameters: 452592

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/6_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	44	3.2	1771	3	US-08-533-669A-7	Sequence 7, Appli
c	2	44	3.2	1771	4	US-08-511-872-1	Sequence 1, Appli
	3	39.6	2.9	1221	7	5212296-16	Patent No. 5212296
	4	39.6	2.9	1879	7	5212296-5	Patent No. 5212296
c	5	38.6	2.8	5173	1	US-08-242-677-1	Sequence 1, Appli
	6	38.4	2.8	49272	2	US-08-614-770A-1	Sequence 1, Appli
c	7	37.8	2.8	3489	4	US-08-728-323A-1	Sequence 1, Appli
	8	37.8	2.8	32207	3	US-08-770-379-20	Sequence 20, Appl
	9	37.4	2.8	3435	2	US-08-366-577-1	Sequence 1, Appli
	10	37.4	2.8	3435	6	PCT-US96-00005-1	Sequence 1, Appli
	11	35.8	2.6	864	1	US-08-396-650-2	Sequence 2, Appli
	12	35.8	2.6	864	2	US-08-768-626-2	Sequence 2, Appli
	13	35.8	2.6	921	1	US-08-396-650-3	Sequence 3, Appli
	14	35.8	2.6	921	1	US-08-396-650-4	Sequence 4, Appli
	15	35.8	2.6	921	2	US-08-768-626-3	Sequence 3, Appli
	16	35.8	2.6	921	2	US-08-768-626-4	Sequence 4, Appli
	17	35.8	2.6	1712	1	US-08-171-299B-1	Sequence 1, Appli
c	18	35.6	2.6	432	2	US-08-642-255-48	Sequence 48, Appl
c	19	35.6	2.6	756	2	US-08-642-255-50	Sequence 50, Appl
	20	35.6	2.6	1335	6	PCT-US91-06532-1	Sequence 1, Appli
	21	35.4	2.6	432	2	US-08-642-255-48	Sequence 48, Appl
	22	35.4	2.6	472	4	US-08-975-316-28	Sequence 28, Appl
	23	35.2	2.6	3234	1	US-08-264-534-31	Sequence 31, Appl
	24	35.2	2.6	3234	2	US-08-083-590A-10	Sequence 10, Appl
	25	35.2	2.6	3234	2	US-08-465-500-31	Sequence 31, Appl
	26	35.2	2.6	3234	3	US-08-346-128-31	Sequence 31, Appl
	27	34.6	2.6	756	2	US-08-642-255-50	Sequence 50, Appl
	28	34.6	2.6	2569	2	US-08-631-607-1	Sequence 1, Appli
	29	34.6	2.6	3684	4	US-08-760-075A-17	Sequence 17, Appl
c	30	34.4	2.5	330	1	US-07-849-389-6	Sequence 6, Appli
c	31	34.4	2.5	2793	2	US-08-209-747-1	Sequence 1, Appli
c	32	34.4	2.5	2793	2	US-08-458-298-1	Sequence 1, Appli
c	33	34.4	2.5	12412	2	US-08-390-878-18	Sequence 18, Appl
	34	34.2	2.5	895	5	US-08-650-766-4	Sequence 4, Appli
	35	34.2	2.5	1171	5	US-08-650-766-5	Sequence 5, Appli
	36	34.2	2.5	1678	5	US-08-650-766-2	Sequence 2, Appli
	37	34.2	2.5	3318	5	US-08-650-766-3	Sequence 3, Appli
	38	34.2	2.5	3385	5	US-08-650-766-1	Sequence 1, Appli
	39	34.2	2.5	4257	3	US-08-690-473-1	Sequence 1, Appli
c	40	34.2	2.5	12001	2	US-08-458-568A-11	Sequence 11, Appl
c	41	34	2.5	1426	3	US-08-284-465-2	Sequence 2, Appli
c	42	34	2.5	1426	3	US-08-284-465-7	Sequence 7, Appli
	43	34	2.5	43280	3	US-08-804-227C-1	Sequence 1, Appli
c	44	33.8	2.5	702	2	US-08-458-568A-3	Sequence 3, Appli
	45	33.8	2.5	884	3	US-08-901-200A-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-08-533-669A-7/c
; Sequence 7, Application US/08533669A
; Patent No. 5834592
; GENERAL INFORMATION:
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF LEISHMANIASIS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/533,669A
; FILING DATE: 22-SEP-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1771 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1698
US-08-533-669A-7

Query Match 3.2%; Score 44; DB 3; Length 1771;
Best Local Similarity 45.8%; Pred. No. 0.0099;
Matches 152; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

Qy 376 cagcttcctccagaggtggcagcaatggcccggctactaggggacctagacaggagcacg 435

```
Db   1050 CCGCGCCCTCGATAGCAGTCGCCCTTTGCACTCCTCGCTGTAGCTCGACGTCGCCGGCCG 991

Qy    436 ttcagaaagttgctgaagtttgtggtcagcagcctgcaggggggaggactgccgagacggt 495

Db    990 CTCATCTTTTTCCTCCTCCTCAGCCTCCAGCTCCGCGCAGAGACGCGCTGCCTCCTCGAG 931

Qy    496 gtgcagcgtcttggggtcagcgccaacctgccggaggagcagctgggtgccctgctggca 555

Db    930 GGCCTGGGAGCGGGCCTGCTCGGCCTCGTGCATGGCCTCCAGGCGGGCCGCCTCCTCGAG 871

Qy    556 ggcatgcacacactgctccagcaggccctccgtctgccccccaccagcctgaagcctgac 615

Db    870 GGCCTGGACACGGGCCTGCTCGCCCTCGTGCATGACATCCAGGCGGGCCGCCTCCTCGGC 811

Qy    616 accttcagggaccagctccaggagctctgcatcccccaagacctggtcggggacttggcc 675

Db    810 CGCCTCCAGCTCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGACGCGGGCCTGCTC 751

Qy    676 agcgtggtatttgggagccagcggcccctcct 707

Db    750 GGCCTCGTGCATGACATCCAGGCGGGCCGCCT 719


RESULT   2
US-08-511-872-1/c
; Sequence 1, Application US/08511872
; Patent No. 5965142
;  GENERAL INFORMATION:
;    APPLICANT:  Dillon, Davin C.
;    APPLICANT:  Reed, Steven G.
;    APPLICANT:  Day, Craig H.
;    TITLE OF INVENTION:  POLYPEPTIDES AND METHODS FOR THE
;    TITLE OF INVENTION:  DETECTION OF L. tropica INFECTION
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  94104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/511,872
;      FILING DATE:  04-AUG-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5965142tenburg, Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  210121.405
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;      TELEX:  3723836 SEEDANDBERRY
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1771 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1698
US-08-511-872-1


  Query Match             3.2%;  Score 44;  DB 4;  Length 1771;
  Best Local Similarity   45.8%;  Pred. No. 0.0099;
  Matches  152;  Conservative    0;  Mismatches  180;  Indels    0;  Gaps    0;

Qy    376 cagcttcctccagaggtggcagcaatggcccggctactaggggacctagacaggagcacg 435

Db   1050 CCGCGCCCTCGATAGCAGTCGCCCTTTGCACTCCTCGCTGTAGCTCGACGTCGCCGGCCG 991

Qy    436 ttcagaaagttgctgaagtttgtggtcagcagcctgcaggggggaggactgccgagacggt 495

Db    990 CTCATCTTTTTCCTCCTCCTCAGCCTCCAGCTCCGCGCAGAGACGCGCTGCCTCCTCGAG 931

Qy    496 gtgcagcgtcttggggtcagcgccaacctgccggaggagcagctgggtgccctgctggca 555

Db    930 GGCCTGGGAGCGGGCCTGCTCGGCCTCGTGCATGGCCTCCAGGCGGGCCGCCTCCTCGAG 871

Qy    556 ggcatgcacacactgctccagcaggccctccgtctgccccccaccagcctgaagcctgac 615

Db    870 GGCCTGGACACGGGCCTGCTCGCCCTCGTGCATGACATCCAGGCGGGCCGCCTCCTCGGC 811

Qy    616 accttcagggaccagctccaggagctctgcatcccccaagacctggtcggggacttggcc 675

Db    810 CGCCTCCAGCTCCGCGCGGAGACGCGCTGCCTCCTCGAGGGCCTGGACGCGGGCCTGCTC 751

Qy    676 agcgtggtatttgggagccagcggcccctcct 707

Db    750 GGCCTCGTGCATGACATCCAGGCGGGCCGCCT 719


RESULT   3
5212296-16
;Patent No. 5212296
;    APPLICANT: DEAN, CAROLINE;HARDER, PATRICIA A.;LETO, KENNETH
;J.; O'KEEFE, DANIEL P.;OMER, CHARLES A.;ROMESSER, JAMES A.
;TEPPERMAN, JAMES M.
;    TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING
;CYTOCHROMES
;    NUMBER OF SEQUENCES: 19
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/569,781
;      FILING DATE: 23-AUG-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 464,499
;      FILING DATE: 12-JAN-1990
;      APPLICATION NUMBER: 405,605
;      FILING DATE: 11-SEP-1989
;SEQ ID NO:16:
;      LENGTH: 1221
5212296-16


  Query Match             2.9%;  Score 39.6;  DB 7;  Length 1221;
  Best Local Similarity   56.0%;  Pred. No. 0.13;
  Matches   75;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy    574 cagcaggccctccgtctgccccccaccagcctgaagcctgacaccttcagggaccagctc 633

Db    796 cagtacgccgccctgcgcgccgaccgcagcctcgtgcccggcgcggtggaggaaactgctc 855

Qy    634 caggagctctgcatcccccaagacctggtcggggacttggccagcgtggtatttgggagc 693

Db    856 cgctacctcgccatcgccgacatcgcgggcggccgcgtcgccacggcggacatcgaggtc 915

Qy    694 cagcggcccctcct 707

Db    916 gaggggcacctcat 929


RESULT   4
5212296-5
;Patent No. 5212296
;    APPLICANT: DEAN, CAROLINE;HARDER, PATRICIA A.;LETO, KENNETH
;J.; O'KEEFE, DANIEL P.;OMER, CHARLES A.;ROMESSER, JAMES A.
;TEPPERMAN, JAMES M.
;    TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING
;CYTOCHROMES
```

```
;    NUMBER OF SEQUENCES: 19
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/569,781
;      FILING DATE: 23-AUG-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 464,499
;      FILING DATE: 12-JAN-1990
;      APPLICATION NUMBER: 405,605
;      FILING DATE: 11-SEP-1989
;SEQ ID NO:5:
;      LENGTH: 1879
5212296-5


  Query Match              2.9%;  Score 39.6;  DB 7;  Length 1879;
  Best Local Similarity   56.0%;  Pred. No. 0.16;
  Matches   75;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy     574 cagcaggccctccgtctgccccccaccagcctgaagcctgacaccttcagggaccagctc 633
           ||| | |||  ||  |   ||  |  ||||||   ||| | | |  |   ||| | ||||
Db     923 cagtacgccgccctgcgcgccgaccgcagcctcgtgcccggcgcggtggaggaactgctc 982

Qy     634 caggagctctgcatcccccaagacctggtcggggacttggccagcgtggtatttgggagc 693
           |   | |||  |||| || |   |  || |||   | | ||||  | ||   | | |  |
Db     983 cgctacctcgccatcgccgacatcgcgggcggccgcgtcgccacggcggacatcgaggtc 1042

Qy     694 cagcggcccctcct 707
             || ||| |||| |
Db    1043 gaggggcacctcat 1056


RESULT   5
US-08-242-677-1/c
: Sequence 1, Application US/08242677
; Patent No. 5677143
;  GENERAL INFORMATION:
;    APPLICANT:  Gaynor, Richard B
;    APPLICANT:  Wu, Foon W.
;    TITLE OF INVENTION:  Cellular Nucleic Acid Binding Protein
;    TITLE OF INVENTION:  and Uses Thereof in regulating Gene Expression and in the
;    TITLE OF INVENTION:  Treatment of AIDS
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  TX
;      COUNTRY:  USA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/242,677
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mayfield, Denise L.
;      REGISTRATION NUMBER:  33,732
;      REFERENCE/DOCKET NUMBER:  UTSD:401
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  713-787-1400
;      TELEFAX:  713-789-2679
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5173 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..4863
US-08-242-677-1


  Query Match              2.8%;  Score 38.6;  DB 1;  Length 5173;
  Best Local Similarity   48.4%;  Pred. No. 0.48;
  Matches  107;  Conservative    0;  Mismatches  114;  Indels    0;  Gaps    0;

Qy     507 tggggtcagcgccaacctgccggaggagcagctgggtgccctgctggcaggcatgcacac 566
           ||  ||| || || | || |     | |  |   |  | |||| |    |    || | |
Db     907 TGCAGTCGGCCCCCAGCTCCGCCGACACCTCCACCGCCCTCTGCAGCAGGTAGCGCGCTC 848

Qy     567 actgctccagcaggccctcctctgccccccaccagcctgaagcctgacaccttcaggga 626
             ||     |  |||| | |||| ||| |||   || |||||  | | |     |     |
Db     847 GCTTGCGCGTCAGGGCGTCCGCCTGGCCCAGCCCCGCCTGCACCGTCCTCCAGAAGCGCC 788

Qy     627 ccagctccaggagctctgcatccccccaagacctggtcggggacttggccagcgtggtatt 686
               | ||   || |   || ||  |  |  | || || |  || |  ||  |   |
Db     787 AGCAGCGCCGGGCGTCCGGGCCCGCCTCGCGCGCGCCGCGGGCGCGGTCGCCGCCGGGCT 728

Qy     687 tgggagccagcggcccctccttgattctgtggcccagcagc 727
             |||  |||| | |  ||  |   ||  || ||||||||
Db     727 CGGGCAACAGCTTCTCGGCCAGGGCGCTCAGGACCAGCAGC 687


RESULT   6
US-08-614-770A-1
; Sequence 1, Application US/08614770A
; Patent No. 5773267
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAM R. JACOBS AND GRAHAM F. HATFULL
;    TITLE OF INVENTION:  D29 SHUTTLE PHASMIDS AND USES THEREOF
;    NUMBER OF SEQUENCES:  1
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  AMSTER, ROTHSTEIN & EBENSTEIN
;      STREET:  90 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  U.S.A.
;      ZIP:  10016
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH 1.44 Mb STORAGE
;      MEDIUM TYPE:  DISKETTE
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCII
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/614,770A
;      FILING DATE:  MARCH 7, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ELIZABETH A. BOGOSIAN
;      REGISTRATION NUMBER:  39,911
;      REFERENCE/DOCKET NUMBER:  96700/402
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 697-5995
;      TELEFAX:  (212) 286-0854 or 286-0082
;      TELEX:  TWX 710-581-4766
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  49272
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  DOUBLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:
;      DESCRIPTION:  OLIGONUCLEOTIDE
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  MYCOBACTERIOPHAGE
;      INDIVIDUAL ISOLATE:  D29
US-08-614-770A-1
```

```
  Query Match             2.8%;  Score 38.4;  DB 2;  Length 49272;
  Best Local Similarity   48.2%;  Pred. No. 1.6;
  Matches  108;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy     523 ctgccggaggagcagctgggtgccctgctggcaggcatgcacacactgctccagcaggcc 582
           ||||  || | |  ||| || |   || | || | | |     | ||   ||||||  |
Db   43074 CTGCGTGAAGCGTTGCTCGGCGGTGTGGTCGCCGACGTACGGGAAGTGGCTCAGCAGCAC 43133

Qy     583 ctccgtctgccccccaccagcctgaagcctgacaccttcagggaccagctccaggagctc 642
             |  || || || || ||  |  |  |   |  |       | | ||  |  || |   ||
Db   43134 GTTGGTGTGACCGCCTCCGTCGAGGGGGACGCGAAGCCTCGCGGCCGTCGACATGTAATC 43193

Qy     643 tgcatcccccaagacctggtcggggacttggccagcgtggtatttgggagccagcggccc 702
               |   | ||  | |   ||  |   ||  |||||     |     |  | ||||
Db   43194 GAGCACGTTCCAGTACAGCCGGGACAGCCGGGGAGCGTCCCGGTACATCGGGTGAGGCCG 43253

Qy     703 ctccttgattctgtggcccagcagcaggggcctggctgccgca 746
             || | | |  ||  |||| ||  | |||||||| |   ||
Db   43254 ATCATGGTTGCCCAGGATCAGCCGCTTGCGGCCTGGCCGGTTCA 43297


RESULT   7
US-08-728-323A-1/c
; Sequence 1, Application US/08728323A
; Patent No. 5948676
;  GENERAL INFORMATION:
;    APPLICANT:  Chang, Yuan
;    APPLICANT:  Bohenzky, Roy A.
;    APPLICANT:  Russo, James J.
;    APPLICANT:  Edelman, Isidore S.
;    APPLICANT:  Moore, Patrick S.
;    TITLE OF INVENTION:  Immediate Early Protein From Kaposi's
;    TITLE OF INVENTION:  Sarcoma-Associated Herpesvirus, DNA
;    TITLE OF INVENTION:  Encoding Same And Uses Thereof
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/728,323A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  0575/52268/JPW/MSC/SKS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-278-0400
;      TELEFAX:  212-391-0525
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3489 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..3489
US-08-728-323A-1

  Query Match             2.8%;  Score 37.8;  DB 4;  Length 3489;
  Best Local Similarity   47.0%;  Pred. No. 0.65;
  Matches  117;  Conservative    0;  Mismatches  132;  Indels    0;  Gaps    0;

Qy     497 tgcagcgtcttggggtcagcgccaacctgccggaggagcagctgggtgccctgctggcag 556
           | | ||  ||   | ||| |      |||| |      | ||||     | | |||
Db    2198 TCCTGCTGCTGCTGCTCATCCTGCTGCTGCTGCTCATCCTGCTGCTGCTCATCCTGCTGC 2139

Qy     557 gcatgcacacactgctccagcaggccctccgtctgccccccaccagcctgaagcctgaca 616
           | ||| ||  ||||| |  | |  | |||  |||| ||     || ||||  ||   ||
Db    2138 TCCTGCTCATCCTGCTGCTCCTGCTCATCCTGCTGCTCCTGCTCATCCTGCTGCTGCTCA 2079

Qy     617 ccttcagggaccagctccaggagctctgcatcccccaagacctggtcggggacttggcca 676
           | |   |   |  ||| |  |||  |||||  |     ||  ||  |    ||  ||
Db    2078 TCCTGCTGCTGCTCATCCTGCTGCTGCTCATCCTGCTGCTGCTCATCCTGCTGCTGCTCA 2019

Qy     677 gcgtggtatttgggagccagcggcccctccttgattctgtggcccagcagcaggggggcct 736
           | || |  |    | || || ||  ||| |        ||  |   | || |  | |
Db    2018 TCCTGCTGCTGCTCATCCTGCTGCTGCTCATCCTGCTGCTGCTCATCCTGCTGCTGCTCA 1959

Qy     737 ggctgccgc 745
             |||| ||
Db    1958 TCCTGCTGC 1950


RESULT   8
US-08-770-379-20
; Sequence 20, Application US/08770379
; Patent No. 5849564
;  GENERAL INFORMATION:
;    APPLICANT:  Chang, Yuan
;    APPLICANT:  Bohenzky, Roy A.
;    APPLICANT:  Russo, James J.
;    APPLICANT:  Edelman, Isidore S.
;    APPLICANT:  Moore, Patrick S.
;    TITLE OF INVENTION:  POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
;    TITLE OF INVENTION:  HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/770,379
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  52342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  32207 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-770-379-20

  Query Match             2.8%;  Score 37.8;  DB 3;  Length 32207;
  Best Local Similarity   47.0%;  Pred. No. 1.9;
  Matches  117;  Conservative    0;  Mismatches  132;  Indels    0;  Gaps    0;
```

```
Qy    497 tgcagcgtcttggggtcagcgccaacctgccggaggagcagctgggtgccctgctggcag 556
          | | ||  ||   | ||| |      |||| |      | ||||     | | |||
Db  19799 TCCTGCTGCTGCTGCTCATCCTGCTGCTGCTGCTCATCCTGCTGCTGCTCATCCTGCTGC 19858

Qy    557 gcatgcacacactgctccagcaggccctccgtctgccccccaccagcctgaagcctgaca 616
          | ||| ||  ||||| |  | |  | |||  |||| ||    || ||||  ||    ||
Db  19859 TCCTGCTCATCCTGCTGCTCCTGCTCATCCTGCTGCTCCTGCTCATCCTGCTGCTGCTCA 19918

Qy    617 ccttcagggaccagctccaggagctctgcatcccccaagacctggtcggggacttggcca 676
          | |   |   |   ||| |  |||   |||||  |     ||  ||  |    ||  ||
Db  19919 TCCTGCTGCTGCTCATCCTGCTGCTGCTCATCCTGCTGCTGCTCATCCTGCTGCTGCTCA 19978

Qy    677 gcgtggtatttgggagccagcggccccctccttgattctgtggcccagcagcaggggcct 736
          | || |  |    | || || ||  ||| |       ||  |  | || |  |  |  |
Db  19979 TCCTGCTGCTGCTCATCCTGCTGCTGCTCATCCTGCTGCTGCTCATCCTGCTGCTGCTCA 20038

Qy    737 ggctgccgc 745
            |||| ||
Db  20039 TCCTGCTGC 20047
```

```
RESULT   9
US-08-366-577-1
; Sequence 1, Application US/08366577
; Patent No. 5728523
;  GENERAL INFORMATION:
;    APPLICANT:  Vogelstein, Bert
;    APPLICANT:  Kinzler, Kenneth W.
;    TITLE OF INVENTION:  POLYMERASE DELTA MUTATIONS IN COLORECTAL
;    TITLE OF INVENTION:  TUMORS WITH REPLICATION ERRORS
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Banner, Birch, McKie & Beckett
;      STREET:  1001 G Street N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20001
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/366,577
;      FILING DATE:  12-DEC-1994
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kagan, Sarah A.
;      REGISTRATION NUMBER:  32,141
;      REFERENCE/DOCKET NUMBER:  01107.48554
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-508-9100
;      TELEFAX:  202-508-9299
;      TELEX:  197430 BBMB UT
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3435 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  43..3364
US-08-366-577-1
```

```
  Query Match             2.8%;  Score 37.4;  DB 2;  Length 3435;
  Best Local Similarity   46.4%;  Pred. No. 0.83;
  Matches  122;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy    430 agcacgttcagaaagttgctgaagtttgtggtcagcagcctgcaggggggaggactgccga 489
          |||| | |  |      ||| || |    || |       |||| | ||||| ||   |
Db   1390 AGCATGGTGGGCCGCGTGCAGATGGACATGCTGCAGGTGCTGCTGCGGGAGTACAAGCTC 1449

Qy    490 gacggtgtgcagcgtcttggggtcagcgccaacctgccggaggagcagctgggtgccctg 549
            |        ||    ||  || |||  | || | | ||  ||||||  ||  | | ||
Db   1450 CGCTCCCACACGCTCAATGCCGTGAGCTTCCACTTCCTGGGCGAGCAGAAGGAGGACGTG 1509

Qy    550 ctggcaggcatgcacacactgctccagcaggccctccgtctgccccccaccagcctgaag 609
          | |    ||||   |||    || ||| | |    |   | | ||| |  | |||||
Db   1510 CAGCACAGCATCATCACCGACCTGCAGAATGGGAACGACCAGACCCGCCGCCGCCTGGCT 1569

Qy    610 cctgacaccttcagggaccagctccaggagctctgcatcccccaagacctggtcggggac 669
            ||  | | |  |||      || |   ||  |  | |   |    ||  | | |
Db   1570 GTGTACTGCCTGAAGGATGCCTACCTGCCACTGCGGCTGCTGGAGCGGCTCATGGTGCTG 1629

Qy    670 ttggccagcgtggtatttgggag 692
            ||  |  |||||   | | |||
Db   1630 GTGAACGCCGTGGAGATGGCGAG 1652
```

```
RESULT  10
PCT-US96-00005-1
; Sequence 1, Application PC/TUS9600005
;  GENERAL INFORMATION:
;    APPLICANT:  Vogelstein, Bert
;    APPLICANT:  Kinzler, Kenneth W.
;    TITLE OF INVENTION:  POLYMERASE DELTA MUTATIONS IN COLORECTAL
;    TITLE OF INVENTION:  TUMORS WITH REPLICATION ERRORS
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Banner & Allegretti, Ltd.
;      STREET:  1001 G Street N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20001
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/00005
;      FILING DATE:  2-JAN-96
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kagan, Sarah A.
;      REGISTRATION NUMBER:  32,141
;      REFERENCE/DOCKET NUMBER:  01107.53505
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-508-9100
;      TELEFAX:  202-508-9299
;      TELEX:  197430 BBMB UT
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3435 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  43..3364
PCT-US96-00005-1
```

```
  Query Match             2.8%;  Score 37.4;  DB 6;  Length 3435;
  Best Local Similarity   46.4%;  Pred. No. 0.83;
  Matches  122;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy     430 agcacgttcagaaagttgctgaagtttgtggtcagcagcctgcaggggggaggactgccga 489
           |||| | |   |     ||| || |    || |       |||| | ||||| ||   |
Db    1390 AGCATGGTGGGCCGCGTGCAGATGGACATGCTGCAGGTGCTGCTGCGGGAGTACAAGCTC 1449

Qy     490 gacggtgtgcagcgtcttggggtcagcgccaacctgccggaggagcagctgggtgccctg 549
           |      ||    ||  || ||| | || | | ||  ||||||  ||  | | ||
Db    1450 CGCTCCCACACGCTCAATGCCGTGAGCTTCCACTTCCTGGGCGAGCAGAAGGAGGACGTG 1509

Qy     550 ctggcaggcatgcacacactgctccagcaggccctccgtctgccccccaccagcctgaag 609
           | |    ||||   |||    || ||| | |    |   | | ||| |  | |||||
Db    1510 CAGCACAGCATCATCACCGACCTGCAGAATGGGAACGACCAGACCCGCCGCCGCCTGGCT 1569

Qy     610 cctgacaccttcagggaccagctccaggagctctgcatcccccaagacctggtcggggac 669
             ||  | | | |||      || |   ||  |  | |   |    ||  | | |
Db    1570 GTGTACTGCCTGAAGGATGCCTACCTGCCACTGCGGCTGCTGGAGCGGCTCATGGTGCTG 1629

Qy     670 ttggccagcgtggtatttgggag 692
           ||  |  |||||   | | |||
Db    1630 GTGAACGCCGTGGAGATGGCGAG 1652


RESULT  11
US-08-396-650-2
; Sequence 2, Application US/08396650
; Patent No. 5629204
;  GENERAL INFORMATION:
;    APPLICANT:  HONJO, TASUKU
;    APPLICANT:  ISHIDA, YASUMASA
;    APPLICANT:  SHINOHARA, TAKASHI
;    TITLE OF INVENTION:  A NOVEL PEPTIDE RELATED TO HUMAN
;    TITLE OF INVENTION:  PROGRAMMED CELL DEATH AND DNA ENCODING IT
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SUGHRUE, MION, ZINN, MACPEAK & SEAS
;      STREET:  2100 Pennsylvania Avenue, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20037-3202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/396,650
;      FILING DATE:  01-MAR-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 55224/1994
;      FILING DATE:  01-MAR-1994
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)293-7060
;      TELEFAX:  (202)293-7860
;      TELEX:  6491103
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  864 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-396-650-2


  Query Match             2.6%;  Score 35.8;  DB 1;  Length 864;
  Best Local Similarity   49.7%;  Pred. No. 1.2;
  Matches   91;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy     457 gtggtcagcagcctgcaggggaggactgccgagacggtgtgcagcgtcttggggtcagc 516
           | || | |||||| ||  ||  ||||||||||   | ||||    |  ||     | |
Db     250 GAGGACCGCAGCCAGCCCGGCCAGGACTGCCGCTTCCGTGTCACACAACTGCCCAACGGG 309

Qy     517 gccaacctgccggaggagcagctgggtgccctgctggcaggcatgcacacactgctccag 576
           || | |    |      | |  | |||| ||     | ||    |||    |||
Db     310 CGTGACTTCCACATGAGCGTGGTCAGGGCCCGGCGCAATGACAGCGGCACCTACCTCTGT 369

Qy     577 caggccctccgtctgccccccaccagcctgaagcctgacaccttcagggaccagctccag 636
           ||||  ||   ||| ||||||     | ||     || | | |  |||   |||||  |
Db     370 GGGGCCATCTCCCTGGCCCCCAAGGCGCAGATCAAAGAGAGCCTGCGGGCAGAGCTCAGG 429

Qy     637 gag 639
           | |
Db     430 GTG 432


RESULT  12
US-08-768-626-2
; Sequence 2, Application US/08768626
; Patent No. 5698520
;  GENERAL INFORMATION:
;    APPLICANT:  HONJO, TASUKU
;    APPLICANT:  ISHIDA, YASUMASA
;    APPLICANT:  SHINOHARA, TAKASHI
;    TITLE OF INVENTION:  A NOVEL PEPTIDE RELATED TO HUMAN
;    TITLE OF INVENTION:  PROGRAMMED CELL DEATH AND DNA ENCODING IT
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SUGHRUE, MION, ZINN, MACPEAK & SEAS
;      STREET:  2100 Pennsylvania Avenue, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20037-3202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/768,626
;      FILING DATE:  01-MAR-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 55224/1994
;      FILING DATE:  01-MAR-1994
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)293-7060
;      TELEFAX:  (202)293-7860
;      TELEX:  6491103
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  864 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-768-626-2


  Query Match             2.6%;  Score 35.8;  DB 2;  Length 864;
  Best Local Similarity   49.7%;  Pred. No. 1.2;
  Matches   91;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy     457 gtggtcagcagcctgcaggggaggactgccgagacggtgtgcagcgtcttggggtcagc 516
           | || | |||||| ||  ||  ||||||||||   | ||||    |  ||     | |
Db     250 GAGGACCGCAGCCAGCCCGGCCAGGACTGCCGCTTCCGTGTCACACAACTGCCCAACGGG 309

Qy     517 gccaacctgccggaggagcagctgggtgccctgctggcaggcatgcacacactgctccag 576
           || | |    |      | |  | |||| ||     | ||    |||    |||
```

```
Db    310 CGTGACTTCCACATGAGCGTGGTCAGGGCCCGGCGCAATGACAGCGGCACCTACCTCTGT 369

Qy    577 caggccctccgtctgccccccaccagcctgaagcctgacaccttcagggaccagctccag 636
          |||| ||   ||| ||||||     | ||     || | | |  |||   |||||  |
Db    370 GGGGCCATCTCCCTGGCCCCCAAGGCGCAGATCAAAGAGAGCCTGCGGGCAGAGCTCAGG 429

Qy    637 gag 639
          | |
Db    430 GTG 432


RESULT  13
US-08-396-650-3
; Sequence 3, Application US/08396650
; Patent No. 5629204
;  GENERAL INFORMATION:
;    APPLICANT:  HONJO, TASUKU
;    APPLICANT:  ISHIDA, YASUMASA
;    APPLICANT:  SHINOHARA, TAKASHI
;    TITLE OF INVENTION:  A NOVEL PEPTIDE RELATED TO HUMAN
;    TITLE OF INVENTION:  PROGRAMMED CELL DEATH AND DNA ENCODING IT
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SUGHRUE, MION, ZINN, MACPEAK & SEAS
;      STREET:  2100 Pennsylvania Avenue, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20037-3202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/396,650
;      FILING DATE:  01-MAR-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 55224/1994
;      FILING DATE:  01-MAR-1994
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)293-7060
;      TELEFAX:  (202)293-7860
;      TELEX:  6491103
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  921 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-396-650-3


  Query Match             2.6%;  Score 35.8;  DB 1;  Length 921;
  Best Local Similarity   49.7%;  Pred. No. 1.2;
  Matches   91;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy    457 gtggtcagcagcctgcaggggggaggactgccgagacggtgtgcagcgtcttggggtcagc 516
          | || | ||||||  ||  ||  ||||||||||    | ||||     |  ||       | |
Db    274 GAGGACCGCAGCCAGCCCGGCCAGGACTGCCGCTTCCGTGTCACACAACTGCCCAACGGG 333

Qy    517 gccaacctgccggaggagcagctgggtgccctgctggcaggcatgcacacactgctccag 576
            || | |     |     | |  | |||| ||     | ||    |||     |||
Db    334 CGTGACTTCCACATGAGCGTGGTCAGGGCCCGGCGCAATGACAGCGGCACCTACCTCTGT 393

Qy    577 caggccctccgtctgccccccaccagcctgaagcctgacaccttcagggaccagctccag 636
            |||| ||   ||| ||||||     | ||     || | | |  |||   |||||  |
Db    394 GGGGCCATCTCCCTGGCCCCCAAGGCGCAGATCAAAGAGAGCCTGCGGGCAGAGCTCAGG 453

Qy    637 gag 639
          | |
Db    454 GTG 456


RESULT  14
US-08-396-650-4
; Sequence 4, Application US/08396650
; Patent No. 5629204
;  GENERAL INFORMATION:
;    APPLICANT:  HONJO, TASUKU
;    APPLICANT:  ISHIDA, YASUMASA
;    APPLICANT:  SHINOHARA, TAKASHI
;    TITLE OF INVENTION:  A NOVEL PEPTIDE RELATED TO HUMAN
;    TITLE OF INVENTION:  PROGRAMMED CELL DEATH AND DNA ENCODING IT
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SUGHRUE, MION, ZINN, MACPEAK & SEAS
;      STREET:  2100 Pennsylvania Avenue, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20037-3202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/396,650
;      FILING DATE:  01-MAR-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 55224/1994
;      FILING DATE:  01-MAR-1994
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)293-7060
;      TELEFAX:  (202)293-7860
;      TELEX:  6491103
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  921 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      CELL LINE:  YTC3
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  25..888
;      IDENTIFICATION METHOD:  by similarity to some other pattern
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  25..84
;      IDENTIFICATION METHOD:  by similarity with known sequence
;      IDENTIFICATION METHOD:  or to an established consensus
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  85..888
;      IDENTIFICATION METHOD:  by similarity with known sequence
;      IDENTIFICATION METHOD:  or to an established consensus
US-08-396-650-4


  Query Match             2.6%;  Score 35.8;  DB 1;  Length 921;
  Best Local Similarity   49.7%;  Pred. No. 1.2;
  Matches   91;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy    457 gtggtcagcagcctgcaggggggaggactgccgagacggtgtgcagcgtcttggggtcagc 516
          | || | ||||||  ||  ||  ||||||||||    | ||||     |  ||       | |
Db    274 GAGGACCGCAGCCAGCCCGGCCAGGACTGCCGCTTCCGTGTCACACAACTGCCCAACGGG 333

Qy    517 gccaacctgccggaggagcagctgggtgccctgctggcaggcatgcacacactgctccag 576
```

```
Db    334 CGTGACTTCCACATGAGCGTGGTCAGGGCCCGGCGCAATGACAGCGGCACCTACCTCTGT 393

Qy    577 caggccctccgtctgccccccaccagcctgaagcctgacaccttcagggaccagctccag 636

Db    394 GGGGCCATCTCCCTGGCCCCCAAGGCGCAGATCAAAGAGAGCCTGCGGGCAGAGCTCAGG 453

Qy    637 gag 639

Db    454 GTG 456


RESULT 15
US-08-768-626-3
; Sequence 3, Application US/08768626
; Patent No. 5698520
;  GENERAL INFORMATION:
;    APPLICANT:  HONJO, TASUKU
;    APPLICANT:  ISHIDA, YASUMASA
;    APPLICANT:  SHINOHARA, TAKASHI
;    TITLE OF INVENTION:  A NOVEL PEPTIDE RELATED TO HUMAN
;    TITLE OF INVENTION:  PROGRAMMED CELL DEATH AND DNA ENCODING IT
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SUGHRUE, MION, ZINN, MACPEAK & SEAS
;      STREET:  2100 Pennsylvania Avenue, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20037-3202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/768,626
;      FILING DATE:  01-MAR-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 55224/1994
;      FILING DATE:  01-MAR-1994
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)293-7060
;      TELEFAX:  (202)293-7860
;      TELEX:  6491103
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  921 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
:    MOLECULE TYPE:  cDNA
US-08-768-626-3


  Query Match             2.6%;  Score 35.8;  DB 2;  Length 921;
  Best Local Similarity   49.7%;  Pred. No. 1.2;
  Matches   91;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy    457 gtggtcagcagcctgcaggggggaggactgccgagacggtgtgcagcgtcttggggtcagc 516

Db    274 GAGGACCGCAGCCAGCCCGGCCAGGACTGCCGCTTCCGTGTCACACAACTGCCCAACGGG 333

Qy    517 gccaacctgccggaggagcagctgggtgccctgctggcaggcatgcacacactgctccag 576

Db    334 CGTGACTTCCACATGAGCGTGGTCAGGGCCCGGCGCAATGACAGCGGCACCTACCTCTGT 393

Qy    577 caggccctccgtctgccccccaccagcctgaagcctgacaccttcagggaccagctccag 636

Db    394 GGGGCCATCTCCCTGGCCCCCAAGGCGCAGATCAAAGAGAGCCTGCGGGCAGAGCTCAGG 453

Qy    637 gag 639

Db    454 GTG 456


Search completed: May 15, 2000, 01:51:08
Job time: 10856 sec
```

OM protein - protein search, using sw model

Run on: May 15, 2000, 02:06:06 ; Search time 38.86 Seconds
(without alignments)
136.533 Million cell updates/sec

Title: US-09-223-796-4
Perfect score: 1132
Sequence: 1 MSAVGAATPYLHHPGDSHSG.....ALVLKEMADLEKRCERRLQD 224

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : A_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	906	80.0	224	1	W37723	Rat Hypertension r
2	93.5	8.3	195	1	W71684	Amino acid sequenc
3	93.5	8.3	196	1	Y07867	Human secreted pro
4	82	7.2	363	1	W06320	Human mitogen-acti
5	82	7.2	363	1	W97670	Human mitogen acti
6	82	7.2	393	1	W06321	Human mitogen-acti
7	82	7.2	393	1	W97671	Human mitogen acti
8	82	7.2	399	1	W06322	Human mitogen-acti
9	82	7.2	399	1	W97672	Human mitogen acti
10	82	7.2	997	1	R15345	Mouse bullous pemp
11	81.5	7.2	384	1	R51884	KAPA synthase enco
12	81.5	7.2	384	1	W73904	E. coli KAP synthe
13	81	7.2	335	1	R99975	Pseudomonas sp. SD
14	81	7.2	383	1	P91395	Peptide expressed
15	79.5	7.0	344	1	R47213	Lipase modulator.
16	76	6.7	696	1	W72087	HSV-2 strain SB5 C
17	75.5	6.7	614	1	W20991	H. pylori inner me
18	75.5	6.7	2629	1	W55885	Rat telomerase. Te
19	75	6.6	506	1	W22305	Rat GADII. Protein
20	75	6.6	506	1	W20064	Rat GADII protein.
21	75	6.6	579	1	W31366	C16N for promoting
22	74	6.5	360	1	W72212	HSV-2 strain SB5 C
23	73.5	6.5	694	1	R04107	DNA-binding protei
24	73	6.4	195	1	P60403	Equine IFN-omega-2
25	73	6.4	493	1	W22306	Human GADII. Prote
26	73	6.4	493	1	W20065	Human GADII protei
27	72.5	6.4	455	1	R35205	Lipase #2. DNA fra
28	72.5	6.4	520	1	W74762	Human secreted pro
29	71.5	6.3	1103	1	R71012	Human neuronal cal
30	71.5	6.3	1103	1	W63151	Human calcium chan
31	71	6.3	169	1	W88390	Human Zneul partia
32	71	6.3	181	1	W88391	Human Zneul partia
33	71	6.3	254	1	W88382	Human neuro-growth
34	71	6.3	273	1	W88381	Human neuro-growth
35	70.5	6.2	2324	1	R05707	Acetyl-CoA-carboxy
36	70	6.2	315	1	W20813	H. pylori secreted
37	70	6.2	579	1	W31365	C16N for promoting
38	70	6.2	1100	1	P95644	Rabbit seletal mus
39	70	6.2	1106	1	W18389	Rabbit calcium cha
40	70	6.2	1106	1	W37712	Rabbit skeletal ca
41	69.5	6.1	211	1	W14495	Urease accessory m
42	69.5	6.1	306	1	R94699	Yeast calcineurin-
43	69.5	6.1	474	1	R14676	Rabbit vitronectin
44	69.5	6.1	784	1	W34179	Human GC binding p
45	69.5	6.1	875	1	W48309	Pisum sativum ACCa

ALIGNMENTS

RESULT 1
W37723
ID W37723 standard; Protein; 224 AA.
AC W37723;
DT 09-JUN-1998 (first entry)
DE Rat Hypertension related calcium regulator.
KW Hypertension related calcium regulated gene; HCaRG; rat parathyroid;
KW extracellular calcium concentration; antibody; hypertension;
KW hyperthyroidism; osteoporosis; heart failure; diabetes; stroke;
KW cancer; inflammatory disease; asthma.
OS Rattus rattus.
FH Key Location/Qualifiers
FT Misc_difference 15. .21
FT /note= "EF-hand like motif"
PN WO9749807-A2.
PD 31-DEC-1997.
PF 23-JUN-1997; CA0439.
PR 21-JUN-1996; US-667495.
PA (GOSS/) GOSSARD F.
PA (HAME/) HAMET P.
PA (LEWA/) LEWANCZUK R.
PA (TREM/) TREMBLAY J.
PI Gossard F, Hamet P, Lewanczuk R, Tremblay J;
DR WPI; 98-077171/07.
DR N-PSDB; V18890.
PT Hypertension related calcium regulated gene - useful to develop
PT products to treat or detect, e.g. hypertension, stroke,
PT osteoporosis, heart failure, cancer, diabetes or asthma
PS Claim 8; Pages 26-27; 46pp; English.
CC This is the amino acid sequence of the hypertension related calcium
CC regulated gene (HCaRG), which was isolated from the rat parathyroid.
CC Its expression is regulated by extracellular calcium concentration.
CC An antibody against the protein, can be used to detect or modulate
CC (e.g. enhance or inhibit) abnormal calcium levels. They can
CC specifically be used to detect or treat, e.g. hypertension,
CC hyperthyroidism, osteoporosis, heart failure, diabetes, stroke, cancer,
CC inflammatory disease, and asthma.
SQ Sequence 224 AA;

Query Match 80.0%; Score 906; DB 1; Length 224;
Best Local Similarity 79.5%; Pred. No. 3.3e-87;
Matches 178; Conservative 23; Mismatches 23; Indels 0; Gaps 0;

```
Qy       1 MSAVGAATPYLHHPGDSHSGRVSFLGAQLPPEVAAMARLLGDLDRSTFRKLLKFVVSSLQ 60
           |||:||| |||||| ||||||||||:|  ||| |:|:|| |||||||||||| || :|
Db       1 MSALGAAAPYLHHPADSHSGRVSFLGSQPSPEVTAVAQLLKDLDRSTFRKLLKLVVGALH 60

Qy      61 GEDCRDGVQRLGVSANLPEEQLGALLAGMHTLLQQALRLPPTSLKPDTFRDQLQELCIPQ 120
           |:|||: |::|| |||| ||:|  |||| |||||||||||| ||||| |:::|||| |||
Db      61 GKDCREAVEQLGASANLSEERLAVLLAGTHTLLQQALRLPPASLKPDAFQEELQELGIPQ 120

Qy     121 DLVGDLASVVFGSQRPLLDSVAQQQGAWLPHVADFRWRVDVAISTSALARSLQPSVLMQL 180
           ||:|||||: |||||||||||||||: |||||: |||||||||||| :||||||||||
Db     121 DLIGDLASLAFGSQRPLLDSVAQQQGSSLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQL 180

Qy     181 KLSDGSAYRFEVPTAKFQEDRYSVALVLKEMADLEKRCERRLQD 224
```

```
          ||:||||:||||| ||||||||||||||||||:|||:|||:|||
Db    181 KLTDGSAHRFEVPIAKFQELRYSVALVLKEMAELEKKCERKLQD 224
```

RESULT 2
W71684
ID W71684 standard; Protein; 195 AA.
AC W71684;
DI 04-DEC-1998 (first entry)
DE Amino acid sequence of the human tumourigenesis associated protein.
KW Human; tumourigenesis associated protein; HTAP; transplantation;
KW tumour; Antagonist; cancer; inflammation; immunological disease;
KW antibody; probe; primer; PCR; amplification; hybridisation;
KW inhibition.
OS Homo sapiens.
PN WO9841635-A1.
PD 24-SEP-1998.
PF 20-MAR-1998; U06066.
PR 20-MAR-1997; US-822260.
PA (INCY-) INCYTE PHARM INC.
PI Goli SK, Hillman JL;
DR WPI; 98-521224/44.
DR N-PSDB; V58281.
PT New tumorigenesis-associated protein and related nucleic acid,
PT vectors, transformed cells - antibodies, agonists and antagonists,
PT for diagnosis, treatment and prevention of abnormal cellular
PT differentiation, particularly cancers and inflammation
PS Claim 1; Fig 1A-1B; 54pp; English.
CC This is the amino acid sequence of the human tumourigenesis
CC associated protein (HTAP), used in the method of the invention. HTAP,
CC is involved with cell proliferation and inflammation. It can be used
CC to stimulate cell proliferation (e.g. of cells intended for
CC transplantation in treatment of tumours or infections, or to treat
CC genetic defects). Antagonists of HTAP are used to treat or prevent
CC a wide range of cancers (adenocarcinoma, melanoma, sarcoma, lymphoma,
CC leukaemia etc.), also inflammation where associated with infection or
CC immunological disease (e.g. asthma, cystic fibrosis, rheumatoid
CC arthritis). HTAP is also used to raise antibodies and to screen
CC libraries for specific-binding agents. The antibodies are used as for
CC diagnosis or monitoring of HTAP-related diseases (in usual
CC immunoassays), in competitive drug screens and to isolate HTAP from
CC its natural sources. HTAP derived probes or primers, are used in
CC standard amplification or hybridisation tests to diagnose HTAP-related
CC diseases; to identify related sequences; for genomic mapping and for
CC screening for specific inhibitors.
SQ Sequence 195 AA;

Query Match 8.3%; Score 93.5; DB 1; Length 195;
Best Local Similarity 22.8%; Pred. No. 0.03;
Matches 33; Conservative 28; Mismatches 57; Indels 27; Gaps 5;

```
Qy      85 LLAGMHTLLQQALRLPPTSLKPDTF-RDQLQELCIPQDLVGDLASVVFGSQRPLLDSVAQ 143
           | || |   :  |    | |:   | |:::: |       :   :: ||
Db      71 LEAGKHRADKSTL---STYLEDCKFDRERIELFCTEYQNNKNSLEILLGS---------- 118

Qy     144 QQGAWLPHVADFRWRVDVAISTSALARSLQPSVLMQLKLSDGSAYRFEVPTAKFQELRYS 203
             |  |||: |  ||::  | |: | |  :|: |: | : :         :  : |: :|
Db     118 -IGRSLPHITDVSWRLEYQIKTNQLHRMYRPAYLVTLSVQN-------TDSPSYPEISFS 169

Qy     204 VAL-----VLKEMADLEKRCERRLQ 223
            ::     :: :: |  |  ||  |
Db     170 CSMEQLQDLVGKLKDASKSLERATQ 194
```

RESULT 3
Y07867
ID Y07867 standard; Protein; 196 AA.
AC Y07867;
DT 06-JUL-1999 (first entry)
DE Human secreted protein fragment encoded from gene 16.
KW Human; secreted protein; treatment; prevention; protein therapy; AIDS;
KW gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
KW developmental abnormality; fetal deficiency; blood disorder; leukemia;
KW immune system disease; autoimmune disease; hepatic disease; lymphoma;
KW renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;
KW cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;
KW pulmonary disorder; transplant rejection; osteoclast; osteoporosis;
KW arthritis; malignancy; digestive; endocrine; infection.
OS Homo sapiens.
PN WO9918208-A1.
PD 15-APR-1999.
PF 01-OCT-1998; U20775.
PR 02-OCT-1997; US-060884.
PR 02-OCT-1997; US-060833.
PR 02-OCT-1997; US-060836.
PR 02-OCT-1997; US-060837.
PR 02-OCT-1997; US-060838.
PR 02-OCT-1997; US-060839.
PR 02-OCT-1997; US-060843.
PR 02-OCT-1997; US-060862.
PR 02-OCT-1997; US-060866.
PR 02-OCT-1997; US-060874.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Carter KC, Duan DR, Endress GA, Feng P, Ferrie AM,
PI Florence KA, Greene JM, Janat F, Lafleur DW, Ni J,
PI Rosen CA, Ruben SM, Shi Y, Young P, Yu G;
DR WPI; 99-264022/22.
DR N-PSDB; X37466.
PT New isolated human genes and the secreted polypeptides they encode
PS Claim 1b; Page 285-286; 368pp; English.
CC This invention describes novel isolated human genes and the secreted
CC proteins they encode. The products of the invention are useful for
CC preventing, treating or ameliorating medical conditions, e.g. by protein
CC or gene therapy. Also pathological conditions can be diagnosed by
CC determining the amount of the new polypeptides in a sample or by
CC determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the 101 polynucleotides, based on
CC which tissues they are most highly expressed in, and include developing
CC products for the diagnosis or treatment of cancer, tumours,
CC neurodegenerative disorders, developmental abnormalities and fetal
CC deficiencies, blood disorders, leukemias, diseases of the immune system,
CC autoimmune diseases, hepatic and renal disease, lymphomas, inflammation,
CC allergies, Alzheimer's and cognitive disorders, schizophrenia, prostate
CC disease, skeletal or cardiac muscle disorders, pulmonary disorders,
CC transplant rejection, disorders involving osteoclasts such as
CC osteoporosis, arthritis or malignancies, digestive/endocrine disorders,
CC infections and AIDS. The human secreted proteins of the invention are
CC represented in Y07852-Y07993 and the encoding nucleic acids are
CC represented in X37451-X37552.
SQ Sequence 196 AA;

Query Match 8.3%; Score 93.5; DB 1; Length 196;
Best Local Similarity 22.8%; Pred. No. 0.03;
Matches 33; Conservative 28; Mismatches 57; Indels 27; Gaps 5;

```
Qy      85 LLAGMHTLLQQALRLPPTSLKPDTF-RDQLQELCIPQDLVGDLASVVFGSQRPLLDSVAQ 143
           | || |   :  |    | |:   | |:::: |       :   :: ||
Db      71 LEAGKHRADKSTL---STYLEDCKFDRERIELFCTEYQNNKNSLEILLGS---------- 118

Qy     144 QQGAWLPHVADFRWRVDVAISTSALARSLQPSVLMQLKLSDGSAYRFEVPTAKFQELRYS 203
             |  |||: |  ||::  | |: | |  :|: |: | : :         :  : |: :|
Db     118 -IGRSLPHITDVSWRLEYQIKTNQLHRMYRPAYLVTLSVQN-------TDSPSYPEISFS 169

Qy     204 VAL-----VLKEMADLEKRCERRLQ 223
            ::     :: :: |  |  ||  |
Db     170 CSMEQLQDLVGKLKDASKSLERATQ 194
```

RESULT 4
W06320
ID W06320 standard; Protein; 363 AA.
AC W06320;
DT 07-FEB-1997 (first entry)
DE Human mitogen-activated protein kinase kinase 4-alpha.

```
KW   Mitogen-activated protein kinase kinase 4-alpha; MKK4-alpha; MAP;
KW   tyrosine kinase; signal transduction; cytokine; oncoprotein;
KW   stress; diagnosis; therapy.
OS   Homo sapiens.
PN   WO9636642-A1.
PD   21-NOV-1996.
PF   26-JAN-1996; U01078.
PR   19-MAY-1995; US-446083.
PR   19-SEP-1995; US-530950.
PA   (DAVI/) DAVIS R J.
PA   (DERI/) DERIJARD B.
PA   (GUPT/) GUPTA S.
PA   (RAIN/) RAINGEAUD J.
PI   Davis RJ,  Derijard B,  Gupta S,  Raingeaud J;
DR   WPI; 97-012035/01.
DR   N-PSDB; T43205.
PT   New mitogen activated protein kinase kinase - useful for treating
PT   ischaemic heart disease, kidney failure etc., also for identifying
PT   modulators for treatment of similar conditions
PS   Claim 11; Fig 6; 104pp; English.
CC   Novel human mitogen activated protein kinase kinase 4-alpha
CC   (MKK4-alpha) (W06320) has serine, threonine and tyrosine kinase
CC   activity, and mediates a signal transduction pathway that activates
CC   human mitogen-activated protein (MAP) kinases p38 and JNK.  It is
CC   an alternatively spliced form of MKK4-gamma (W06322) and MKK4-beta
CC   (W06321) isoforms, which differ slightly at the NH2-terminus.  MKK4
CC   is expressed in all tissues tested, but partic. in skeletal
CC   muscle.  Recombinant MKK4 can be produced in transformed host cells.
CC   MKKs (W06318-22) can be used in the treatment of MKK-related
CC   disorders, e.g. ischaemic heart disease and kidney failure, to
CC   identify modulators of MKK activity, and to raise antibodies.
SQ   Sequence   363 AA;
```

```
  Query Match             7.2%;  Score 82;  DB 1;  Length 363;
  Best Local Similarity   20.2%;  Pred. No. 1.1;
  Matches   59;  Conservative   44;  Mismatches   89;  Indels  100;  Gaps   13;

Qy       3 AVGAATPYLHHPGDSHSGRVSFLGAQLPPEVAAMARLLGDLDRSTFRKLLKFVVSSLQGE 62
           | |:    :| |    ||           :: |: |:   :|    ::|| :   ::
Db      76 AYGSVNKMVHKP----SG-----------QIMAVKRIRSTVDEKEQKQLLMDLDVVMRSS 120

Qy      63 DC-----------RDG-----VQRLGVSAN------------LPEEQLGAL-LAGMHTL 92
           ||           |:|     :: :  | :             :||| || : || :  |
Db     121 DCPYIVQFYGALFREGDCWICMELMSTSFDKFYKYVYSVLDDVIPEEILGKITLATVKAL 180

Qy      93 --LQQALRLPPTSLKPDTFRDQLQELCIPQDLVGDLASVVFGSQRPLLDSVAQQQGAWL- 150
             |:: |::    :||          | |  |::    ||    |:||:|: : |
Db     181 NHLKENLKIIHRDIKPSN---------ILLDRSGNIKLCDFGISGQLVDSIAKTRDAGCR 231

Qy     150 PHVADFR---------------WRVDVAISTSALARSLQP---SVLMQL---------K 181
           |::|  |               | : : :   |  |   |   ||  ||         :
Db     232 PYMAPERIDPSASRQGYDVRSDVWSLGITLYELATGRFPYPKWNSVFDQLTQVVKGDPPQ 291

Qy     182 LSDGSAYRF--------------EVPTAKFQELRYSVALVLKEMADLEKRC 218
           ||:    |               |    |::||     ::: |   :|  |
Db     292 LSNSEEREFSPSFINFVNLCLTKDESKRPKYKELLKHPFILMYEERAVEVAC 343
```

```
RESULT   5
W97670
ID   W97670 standard; Protein; 363 AA.
AC   W97670;
DT   10-MAY-1999  (first entry)
DE   Human mitogen activated protein kinase kinase MKK4 alpha.
KW   MKK4 alpha; human; mitogen activated protein kinase kinase;
KW   MAP kinase kinase; signal transduction; inflammation; psoriasis;
KW   AIDS; cancer; apoptosis; therapy.
OS   Homo sapiens.
PN   WO9902547-A1.
PD   21-JAN-1999.
PF   07-JUL-1998; U14101.
PR   07-JUL-1997; US-888429.
PA   (UYMA-) UNIV MASSACHUSETTS.
PI   Davis RJ, Tournier C, Whitmarsh A;
DR   WPI; 99-120771/10.
DR   N-PSDB; X07067.
PT   New isolated mitogen-activated protein kinase kinase isoforms - used
PT   to develop products for treating e.g. inflammatory disorders,
PT   oxidative damage, proliferative disorders or autoimmune disorders
PS   Example 1; Page 132-133; 168pp; English.
CC   This polypeptide comprises human mitogen activated protein (MAP)
CC   kinase kinase 4 alpha (MKK4 alpha).  Its sequence was deduced from
CC   a cDNA clone (see X07067) derived from human brain mRNA.  MKK4
CC   includes 3 isoforms, termed MKK4 alpha, beta (see W97671) and gamma
CC   (see W97672), that vary slightly at the N-terminus.  MKK4 is a p38
CC   MAP kinase kinase that also phosphorylates the JNK subgroup of MAP
CC   kinases.  MKK3, MKK4, MKK6 and MKK7 (see W97664 and W97668-72),
CC   described in the invention, mediate the transduction of specific
CC   signals from the cell surface to the nucleus along specific
CC   pathways.  They are useful for screening reagents which modulate
CC   MKK activity.  Such agents can be used to prevent or treat
CC   MKK-mediated disorders, e.g. inflammation, oxidative damage or
CC   stress-related proliferative disorders, e.g. psoriasis, AIDS,
CC   malignancies of e.g. the skin, bone marrow, lung, liver, breast,
CC   gastrointestinal system and genito-urinary tract.  Agents which
CC   inhibit the activity or expression of MKK inhibit cell growth or
CC   cause apoptosis.  MKK7 polynucleotides (see X07059-64) and
CC   polypeptides (see W97662-67) are claimed.
SQ   Sequence   363 AA;
```

```
  Query Match             7.2%;  Score 82;  DB 1;  Length 363;
  Best Local Similarity   20.2%;  Pred. No. 1.1;
  Matches   59;  Conservative   44;  Mismatches   89;  Indels  100;  Gaps   13;

Qy       3 AVGAATPYLHHPGDSHSGRVSFLGAQLPPEVAAMARLLGDLDRSTFRKLLKFVVSSLQGE 62
           | |:    :| |    ||           :: |: |:   :|    ::|| :   ::
Db      76 AYGSVNKMVHKP----SG-----------QIMAVKRIRSTVDEKEQKQLLMDLDVVMRSS 120

Qy      63 DC-----------RDG-----VQRLGVSAN------------LPEEQLGAL-LAGMHTL 92
           ||           |:|     :: :  | :             :||| || : || :  |
Db     121 DCPYIVQFYGALFREGDCWICMELMSTSFDKFYKYVYSVLDDVIPEEILGKITLATVKAL 180

Qy      93 --LQQALRLPPTSLKPDTFRDQLQELCIPQDLVGDLASVVFGSQRPLLDSVAQQQGAWL- 150
             |:: |::    :||          | |  |::    ||    |:||:|: : |
Db     181 NHLKENLKIIHRDIKPSN---------ILLDRSGNIKLCDFGISGQLVDSIAKTRDAGCR 231

Qy     150 PHVADFR---------------WRVDVAISTSALARSLQP---SVLMQL---------K 181
           |::|  |               | : : :   |  |   |   ||  ||         :
Db     232 PYMAPERIDPSASRQGYDVRSDVWSLGITLYELATGRFPYPKWNSVFDQLTQVVKGDPPQ 291

Qy     182 LSDGSAYRF--------------EVPTAKFQELRYSVALVLKEMADLEKRC 218
           ||:    |               |    |::||     ::: |   :|  |
Db     292 LSNSEEREFSPSFINFVNLCLTKDESKRPKYKELLKHPFILMYEERAVEVAC 343
```

```
RESULT   6
W06321
ID   W06321 standard; Protein; 393 AA.
AC   W06321;
DT   07-FEB-1997  (first entry)
DE   Human mitogen-activated protein kinase kinase 4-beta.
KW   Mitogen-activated protein kinase kinase 4-beta; MKK4-beta; MAP;
KW   tyrosine kinase; signal transduction; cytokine; oncoprotein;
KW   stress; diagnosis; therapy.
OS   Homo sapiens.
PN   WO9636642-A1.
PD   21-NOV-1996.
PF   26-JAN-1996; U01078.
PR   19-MAY-1995; US-446083.
PR   19-SEP-1995; US-530950.
PA   (DAVI/) DAVIS R J.
PA   (DERI/) DERIJARD B.
PA   (GUPT/) GUPTA S.
PA   (RAIN/) RAINGEAUD J.
```

```
PI   Davis RJ,  Derijard B,  Gupta S,  Raingeaud J;
DR   WPI; 97-012035/01.
DR   N-PSDB; T43206.
PT   New mitogen activated protein kinase kinase - useful for treating
PT   ischaemic heart disease, kidney failure etc., also for identifying
PT   modulators for treatment of similar conditions
PS   Claim 15; Fig 7; 104pp; English.
CC   Novel human mitogen activated protein kinase kinase 4-beta
CC   (MKK4-beta) (W06320) has serine, threonine and tyrosine kinase
CC   activity, and mediates a signal transduction pathway that activates
CC   human mitogen-activated protein (MAP) kinases p38 and JNK.  It is
CC   an alternatively spliced form of MKK4-gamma (W06322) and MKK4-alpha
CC   (W06320) isoforms, which differ slightly at the NH2-terminus.  MKK4
CC   is expressed in all tissues tested, but partic. in skeletal
CC   muscle.  Recombinant MKK4 can be produced in transformed host cells.
CC   MKKs (W06318-22) can be used in the treatment of MKK-related
CC   disorders, e.g. ischaemic heart disease and kidney failure, to
CC   identify modulators of MKK activity, and to raise antibodies.
SQ   Sequence   393 AA;


  Query Match             7.2%;  Score 82;  DB 1;  Length 393;
  Best Local Similarity   20.2%;  Pred. No. 1.2;
  Matches   59;  Conservative   44;  Mismatches   89;  Indels  100;  Gaps   13;

Qy       3 AVGAATPYLHHPGDSHSGRVSFLGAQLPPEVAAMARLLGDLDRSTFRKLLKFVVSSLQGE 62
           | |:    :| |    ||          :: |: |:   :|    ::|| :   ::
Db     106 AYGSVNKMVHKP----SG----------QIMAVKRIRSTVDEKEQKQLLMDLDVVMRSS 150

Qy      63 DC----------RDG-----VQRLGVSAN------------LPEEQLGAL-LAGMHTL 92
           ||          |:|     :: :  | :            :||| || : || :  |
Db     151 DCPYIVQFYGALFREGDCWICMELMSTSFDKFYKYVYSVLDDVIPEEILGKITLATVKAL 210

Qy      93 --LQQALRLPPTSLKPDTFRDQLQELCIPQDLVGDLASVVFGSQRPLLDSVAQQQGAWL- 150
             |:: |::    :||           |  |  |::    ||    |:||:|: : |
Db     211 NHLKENLKIIHRDIKPSN---------ILLDRSGNIKLCDFGISGQLVDSIAKTRDAGCR 261

Qy     150 PHVADFR---------------WRVDVAISTSALARSLQP---SVLMQL---------K 181
           |::|  |               | : : :   |  |   |   ||  ||         :
Db     262 PYMAPERIDPSASRQGYDVRSDVWSLGITLYELATGRFPYPKWNSVFDQLTQVVKGDPPQ 321

Qy     182 LSDGSAYRF--------------EVPTAKFQELRYSVALVLKEMADLEKRC 218
           ||:    |               |    |::||     ::: |   :|  |
Db     322 LSNSEEREFSPSFINFVNLCLTKDESKRPKYKELLKHPFILMYEERAVEVAC 373


RESULT   7
W97671
ID   W97671 standard; Protein; 393 AA.
AC   W97671;
DT   10-MAY-1999  (first entry)
DE   Human mitogen activated protein kinase kinase MKK4 beta.
KW   MKK4 beta; human; mitogen activated protein kinase kinase;
KW   MAP kinase kinase; signal transduction; inflammation; psoriasis;
KW   AIDS; cancer; apoptosis; therapy.
OS   Homo sapiens.
PN   WO9902547-A1.
PD   21-JAN-1999.
PF   07-JUL-1998; U14101.
PR   07-JUL-1997; US-888429.
PA   (UYMA-) UNIV MASSACHUSETTS.
PI   Davis RJ, Tournier C, Whitmarsh A;
DR   WPI; 99-120771/10.
DR   N-PSDB; X07068.
PT   New isolated mitogen-activated protein kinase kinase isoforms - used
PT   to develop products for treating e.g. inflammatory disorders,
PT   oxidative damage, proliferative disorders or autoimmune disorders
PS   Example 1; Page 135-136; 168pp; English.
CC   This polypeptide comprises human mitogen activated protein (MAP)
CC   kinase kinase 4 beta (MKK4 beta).  Its sequence was deduced from
CC   a cDNA clone (see X07068) derived from human brain mRNA.  MKK4
CC   includes 3 isoforms, termed MKK4 alpha (see W97670), beta and gamma
CC   (see W97672), that vary slightly at the N-terminus.  MKK4 is a p38
CC   MAP kinase kinase that also phosphorylates the JNK subgroup of MAP
CC   kinases.  MKK3, MKK4, MKK6 and MKK7 (see W97664 and W97668-72),
CC   described in the invention, mediate the transduction of specific
CC   signals from the cell surface to the nucleus along specific
CC   pathways.  They are useful for screening reagents which modulate
CC   MKK activity.  Such agents can be used to prevent or treat
CC   MKK-mediated disorders, e.g. inflammation, oxidative damage or
CC   stress-related proliferative disorders, e.g. psoriasis, AIDS,
CC   malignancies of e.g. the skin, bone marrow, lung, liver, breast,
CC   gastrointestinal system and genito-urinary tract.  Agents which
CC   inhibit the activity or expression of MKK inhibit cell growth or
CC   cause apoptosis.  MKK7 polynucleotides (see X07059-64) and
CC   polypeptides (see W97662-67) are claimed.
SQ   Sequence   393 AA;


  Query Match             7.2%;  Score 82;  DB 1;  Length 393;
  Best Local Similarity   20.2%;  Pred. No. 1.2;
  Matches   59;  Conservative   44;  Mismatches   89;  Indels  100;  Gaps   13;

Qy       3 AVGAATPYLHHPGDSHSGRVSFLGAQLPPEVAAMARLLGDLDRSTFRKLLKFVVSSLQGE 62
           | |:    :| |    ||          :: |: |:   :|    ::|| :   ::
Db     106 AYGSVNKMVHKP----SG----------QIMAVKRIRSTVDEKEQKQLLMDLDVVMRSS 150

Qy      63 DC----------RDG-----VQRLGVSAN------------LPEEQLGAL-LAGMHTL 92
           ||          |:|     :: :  | :            :||| || : || :  |
Db     151 DCPYIVQFYGALFREGDCWICMELMSTSFDKFYKYVYSVLDDVIPEEILGKITLATVKAL 210

Qy      93 --LQQALRLPPTSLKPDTFRDQLQELCIPQDLVGDLASVVFGSQRPLLDSVAQQQGAWL- 150
             |:: |::    :||           |  |  |::    ||    |:||:|: : |
Db     211 NHLKENLKIIHRDIKPSN---------ILLDRSGNIKLCDFGISGQLVDSIAKTRDAGCR 261

Qy     150 PHVADFR---------------WRVDVAISTSALARSLQP---SVLMQL---------K 181
           |::|  |               | : : :   |  |   |   ||  ||         :
Db     262 PYMAPERIDPSASRQGYDVRSDVWSLGITLYELATGRFPYPKWNSVFDQLTQVVKGDPPQ 321

Qy     182 LSDGSAYRF--------------EVPTAKFQELRYSVALVLKEMADLEKRC 218
           ||:    |               |    |::||     ::: |   :|  |
Db     322 LSNSEEREFSPSFINFVNLCLTKDESKRPKYKELLKHPFILMYEERAVEVAC 373


RESULT   8
W06322
ID   W06322 standard; Protein; 399 AA.
AC   W06322;
DT   07-FEB-1997  (first entry)
DE   Human mitogen-activated protein kinase kinase 4-gamma.
KW   Mitogen-activated protein kinase kinase 4-gamma; MKK4-gamma; MAP;
KW   tyrosine kinase; signal transduction; cytokine; oncoprotein;
KW   stress; diagnosis; therapy.
OS   Homo sapiens.
PN   WO9636642-A1.
PD   21-NOV-1996.
PF   26-JAN-1996; U01078.
PR   19-MAY-1995; US-446083.
PR   19-SEP-1995; US-530950.
PA   (DAVI/) DAVIS R J.
PA   (DERI/) DERIJARD B.
PA   (GUPT/) GUPTA S.
PA   (RAIN/) RAINGEAUD J.
PI   Davis RJ,  Derijard B,  Gupta S,  Raingeaud J;
DR   WPI; 97-012035/01.
DR   N-PSDB; T43207.
PT   New mitogen activated protein kinase kinase - useful for treating
PT   ischaemic heart disease, kidney failure etc., also for identifying
PT   modulators for treatment of similar conditions
PS   Claim 19; Fig 8; 104pp; English.
CC   Novel human mitogen activated protein kinase kinase 4-gamma
CC   (MKK4-gamma) (W06322) has serine, threonine and tyrosine kinase
CC   activity, and mediates a signal transduction pathway that activates
CC   human mitogen-activated protein (MAP) kinases p38 and JNK.  Its
CC   amino acid sequence was deduced from a cDNA clone (T43206) obtd.
CC   from a human foetal brain library.  Alternative splicing of the
```

CC sequence yields MKK4 isoforms alpha (W06320) and beta (W06321).
CC MKK4 is expressed in all tissues tested, but partic. in skeletal
CC muscle. Recombinant MKK4 can be produced in transformed host cells.
CC MKKs (W06318-22) can be used in the treatment of MKK-related
CC disorders, e.g. ischaemic heart disease and kidney failure, to
CC identify modulators of MKK activity, and to raise antibodies.
SQ Sequence 399 AA;

Query Match 7.2%; Score 82; DB 1; Length 399;
Best Local Similarity 20.2%; Pred. No. 1.2;
Matches 59; Conservative 44; Mismatches 89; Indels 100; Gaps 13;

Qy 3 AVGAATPYLHHPGDSHSGRVSFLGAQLPPEVAAMARLLGDLDRSTFRKLLKFVVSSLQGE 62
Db 112 AYGSVNKMVHKP----SG----------QIMAVKRIRSTVDEKEQKQLLMDLDVVMRSS 156

Qy 63 DC----------RDG-----VQRLGVSAN----------LPEEQLGAL-LAGMHTL 92
Db 157 DCPYIVQFYGALFREGDCWICMELMSTSFDKFYKYVYSVLDDVIPEEILGKITLATVKAL 216

Qy 93 --LQQALRLPPTSLKPDTFRDQLQELCIPQDLVGDLASVVFGSQRPLLDSVAQQQGAWL- 150
Db 217 NHLKENLKIIHRDIKPSN---------ILLDRSGNIKLCDFGISGQLVDSIAKTRDAGCR 267

Qy 150 PHVADFR----------WRVDVAISTSALARSLQP---SVLMQL---------K 181
Db 268 PYMAPERIDPSASRQGYDVRSDVWSLGITLYELATGRFPYPKWNSVFDQLTQVVKGDPPQ 327

Qy 182 LSDGSAYRF----------EVPTAKFQELRYSVALVLKEMADLEKRC 218
Db 328 LSNSEEREFSPSFINFVNLCLTKDESKRPKYKELLKHPFILMYEERAVEVAC 379

RESULT 9
W97672
ID W97672 standard; Protein; 399 AA.
AC W97672;
DT 10-MAY-1999 (first entry)
DE Human mitogen activated protein kinase kinase MKK4 gamma.
KW MKK4 gamma; human; mitogen activated protein kinase kinase;
KW MAP kinase kinase; signal transduction; inflammation; psoriasis;
KW AIDS; cancer; apoptosis; therapy.
OS Homo sapiens.
PN WO9902547-A1.
PD 21-JAN-1999.
PF 07-JUL-1998; U14101.
PR 07-JUL-1997; US-888429.
PA (UYMA-) UNIV MASSACHUSETTS.
PI Davis RJ, Tournier C, Whitmarsh A;
DR WPI; 99-120771/10.
DR N-PSDB; X07069.
PT New isolated mitogen-activated protein kinase kinase isoforms - used
PT to develop products for treating e.g. inflammatory disorders,
PT oxidative damage, proliferative disorders or autoimmune disorders
PS Example 1; Page 139; 168pp; English.
CC This polypeptide comprises human mitogen activated protein (MAP)
CC kinase kinase 4 gamma (MKK4 gamma). Its sequence was deduced from
CC a cDNA clone (see X07067) derived from human brain mRNA. MKK4
CC includes 3 isoforms, termed MKK4 alpha (see W97670), beta (see
CC W97671) and gamma, that vary slightly at the N-terminus. MKK4 is a
CC p38 MAP kinase kinase that also phosphorylates the JNK subgroup of
CC MAP kinases. MKK3, MKK4, MKK6 and MKK7 (see W97664 and W97668-72),
CC described in the invention, mediate the transduction of specific
CC signals from the cell surface to the nucleus along specific
CC pathways. They are useful for screening reagents which modulate
CC MKK activity. Such agents can be used to prevent or treat
CC MKK-mediated disorders, e.g. inflammation, oxidative damage or
CC stress-related proliferative disorders, e.g. psoriasis, AIDS,
CC malignancies of e.g. the skin, bone marrow, lung, liver, breast,
CC gastrointestinal system and genito-urinary tract. Agents which
CC inhibit the activity or expression of MKK inhibit cell growth or
CC cause apoptosis. MKK7 polynucleotides (see X07059-64) and
CC polypeptides (see W97662-67) are claimed.
SQ Sequence 399 AA;

Query Match 7.2%; Score 82; DB 1; Length 399;
Best Local Similarity 20.2%; Pred. No. 1.2;
Matches 59; Conservative 44; Mismatches 89; Indels 100; Gaps 13;

Qy 3 AVGAATPYLHHPGDSHSGRVSFLGAQLPPEVAAMARLLGDLDRSTFRKLLKFVVSSLQGE 62
Db 112 AYGSVNKMVHKP----SG----------QIMAVKRIRSTVDEKEQKQLLMDLDVVMRSS 156

Qy 63 DC----------RDG-----VQRLGVSAN----------LPEEQLGAL-LAGMHTL 92
Db 157 DCPYIVQFYGALFREGDCWICMELMSTSFDKFYKYVYSVLDDVIPEEILGKITLATVKAL 216

Qy 93 --LQQALRLPPTSLKPDTFRDQLQELCIPQDLVGDLASVVFGSQRPLLDSVAQQQGAWL- 150
Db 217 NHLKENLKIIHRDIKPSN---------ILLDRSGNIKLCDFGISGQLVDSIAKTRDAGCR 267

Qy 150 PHVADFR----------WRVDVAISTSALARSLQP---SVLMQL---------K 181
Db 268 PYMAPERIDPSASRQGYDVRSDVWSLGITLYELATGRFPYPKWNSVFDQLTQVVKGDPPQ 327

Qy 182 LSDGSAYRF----------EVPTAKFQELRYSVALVLKEMADLEKRC 218
Db 328 LSNSEEREFSPSFINFVNLCLTKDESKRPKYKELLKHPFILMYEERAVEVAC 379

RESULT 10
R15345
ID R15345 standard; Protein; 997 AA.
AC R15345;
DT 25-MAR-1992 (first entry)
DE Mouse bullous pemphigoid antigen protein.
KW dermatological syndrome; blisters; murine.
OS Mus musculus.
PN J03206885-A.
PD 10-SEP-1991.
PF 09-APR-1990; 093585.
PR 31-OCT-1989; JP-283947.
PR 09-APR-1990; JP-093585.
PA (TOFU) TONEN CORP.
DR WPI; 91-356566/49.
DR N-PSDB; Q15019.
PT DNA which encodes mouse derived bullous pemphigoid antigen protein -
PT contg. base sequence which encodes polypeptide composed of specified
PT 996 aminoacid sequence
PS Disclosure; Fig 1; 23pp; Japanese.
CC The polypeptide is recognised by auto-antibodies in the serum of a
CC bullous pemphigoid patient and by a human anti-basement membrane zone
CC monoclonal antibody. It can be used for diagnosis of bullous
CC pemphigoid. See also Q15020.
SQ Sequence 997 AA;

Query Match 7.2%; Score 82; DB 1; Length 997;
Best Local Similarity 22.4%; Pred. No. 4.2;
Matches 59; Conservative 30; Mismatches 106; Indels 68; Gaps 11;

Qy 2 SAVGAATPYLHHPGDSHSGRVSFL----------GAQLPPEVAA 35
Db 228 SAFDTAGRECHHPAEISPGNSGHLNLKTRLPLSRWTQEPHQTEGKWPHRAAEQLPKEV-- 286

Qy 36 MARLLG-DLDRSTFRKLLKFVVSSLQGE-----DCRDGVQRLGVSANLPEEQLGALLAGM 89
Db 286 QFRQPGAPLDRESSQPCYSEYFSQTSTELQITFDDKNPITR----------LSEL 330

Qy 90 HTLLQQALR--LPPTSLKPDTFRDQLQELCIPQDLV-----GDLASVVFGSQRPLLDSVA 142
Db 331 ETMREQALHPSRPPVTYQDDKLERELVKLLTPLEIAKNKQCGMHTEVTTLKQEKRLGSSA 390

Qy 143 QQQGAWLPHVADFRWRVDVAISTSALARSLQPSVLMQLKLSDGSAYRFEVPTAKFQELRY 202

```
Db    391 ---GGWMLEGC----RTSGGLKGDFLKKSVEPEASPSLDLNQACSVRDE--EFQFQGLRH 441

Qy    203 SVA---LVLKEMADLEKRCERRL 222
Db    442 TVTGRQLVEAKLLDMRTVEQLRL 464
```

RESULT 11
R51884
ID R51884 standard; Protein; 384 AA.
AC R51884;
DT 17-NOV-1994 (first entry)
DE KAPA synthase encoded by bioF gene in plasmid pB030A-15/9.
KW Biotin; expression; enterobacteria; vitamin H; synthesis;
KW plasmid; pB030A-15/9; bioB; bioF; bioC; bioD; bioA;
KW promoter ptac; biotin synthase; KAPA synthase;
KW 8-amino-7-oxononanoate synthase; pimeloyl-CoA; DTB synthase;
KW dethiobiotin synthase; DAPA synthase;
KW S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase;
KW seborrhoea; dermatitis.
OS Escherichia coli DSM498.
PN WO9408023-A.
PD 14-APR-1994.
PF 01-OCT-1993; E02688.
PR 02-OCT-1992; CH-003124.
PR 15-JUL-1993; CH-002134.
PA (LONZ) LONZA AG.
PI Birch O, Brass J, Fuhrmann M, Shaw N;
DR WPI; 94-135587/16.
DR N-PSDB; Q62386.
PT Biotechnological biotin prodn. using enterobacterial biotin-gene
PT - providing vitamin H in high yield
PS Disclosure; Fig 6, Page 47-55 and 60-65; 92pp; German.
CC Plasmid pB030A-15/9 contains the bioB, bioF, bioC, bioD and bioA
CC genes responsible for biosynthesis of biotin, arranged in a
CC transcription unit. Microorganisms contg. these DNA fragments or
CC plasmids may be used in the prodn. of biotin. Biotin (Vitamin H)
CC may prevent seborrhoea, dermatitis, loss of appetite and tiredness.
SQ Sequence 384 AA;

Query Match 7.2%; Score 81.5; DB 1; Length 384;
Best Local Similarity 19.9%; Pred. No. 1.3;
Matches 56; Conservative 32; Mismatches 90; Indels 103; Gaps 12;

```
Qy      6 AATPYL---HHP--------------GDSHSGRVSFLGAQLPPEVAAMARLLGDLDRST 47
Db     45 SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLG------ 99

Qy     48 FRKLLKF---------VVSSLQGEDCRDGVQRLGVSANLPEEQLGALLAGMHTLLQQALR 98
Db     99 YSRALLFISGFAANQAVIAAMMAKEDRIAADRLS----------------HASLLEAAS 141

Qy     99 LPPTSLKPDTFRD------QLQELCIPQDLVGDLASVVF---GSQRPL--LDSVAQQQGA 147
Db    142 LSPSQLRRFAHNDVTHLARLLASPCPGQQMV--VTEGVFSMDGDSAPLAEIQQVTQQHNG 199

Qy    148 WLPHVADFRWRVDVAISTSALA---------RSLQPSVLM------------------- 179
Db    200 WL--------MVDDAHGTGVIGEQGRGSCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTV 251

Qy    179 ---QLKLSDGSAYRFEVPTAKFQELRYSVALVLKEMADLEK 216
Db    252 ADYLLQFARHLIYSTSMPPAQAQALRASLAVIRSDEGDARR 292
```

RESULT 12
W73904
ID W73904 standard; Protein; 384 AA.
AC W73904;
DT 12-APR-1999 (first entry)
DE E. coli KAP synthetase.
KW DAP aminotransferase; diaminopelargonic acid; transgenic plant;
KW biotin synthase; biotin production; vitamin H; KAP synthetase.
OS Escherichia coli.
PN US5869719-A.
PD 09-FEB-1999.
PF 30-APR-1997; 846338.
PR 30-APR-1997; US-846338.
PR 08-MAR-1995; US-401068.
PA (NOVS) NOVARTIS FINANCE CORP.
PI Patton DA;
DR WPI; 99-152902/13.
DR N-PSDB; X01301.
PT Transgenic plants with high biotin levels - transformed with DNA
PT encoding di:amino-pelargonic acid amino-transferase or biotin
PT synthase
PS Example 2; Column 23-28; 34pp; English.
CC This sequence is the E. coli 7-keto-8-aminopelargonic acid (KAP)
CC synthetase. The gene can be used in the transgenic plant of the
CC invention. The transgenic plant, plant cell or plant tissue is
CC transformed with a chimeric gene encoding diaminopelargonic acid (DAP)
CC aminotransferase or biotin synthase and produces more biotin than a
CC non-transgenic plant, cell or tissue. The plant is used as an improved
CC dietary source of biotin (vitamin H) for humans or animals.
SQ Sequence 384 AA;

Query Match 7.2%; Score 81.5; DB 1; Length 384;
Best Local Similarity 19.9%; Pred. No. 1.3;
Matches 56; Conservative 32; Mismatches 90; Indels 103; Gaps 12;

```
Qy      6 AATPYL---HHP--------------GDSHSGRVSFLGAQLPPEVAAMARLLGDLDRST 47
Db     45 SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLG------ 99

Qy     48 FRKLLKF---------VVSSLQGEDCRDGVQRLGVSANLPEEQLGALLAGMHTLLQQALR 98
Db     99 YSRALLFISGFAANQAVIAAMMAKEDRIAADRLS----------------HASLLEAAS 141

Qy     99 LPPTSLKPDTFRD------QLQELCIPQDLVGDLASVVF---GSQRPL--LDSVAQQQGA 147
Db    142 LSPSQLRRFAHNDVTHLARLLASPCPGQQMV--VTEGVFSMDGDSAPLAEIQQVTQQHNG 199

Qy    148 WLPHVADFRWRVDVAISTSALA---------RSLQPSVLM------------------- 179
Db    200 WL--------MVDDAHGTGVIGEQGRGSCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTV 251

Qy    179 ---QLKLSDGSAYRFEVPTAKFQELRYSVALVLKEMADLEK 216
Db    252 ADYLLQFARHLIYSTSMPPAQAQALRASLAVIRSDEGDARR 292
```

RESULT 13
R99975
ID R99975 standard; Protein; 335 AA.
AC R99975;
DT 29-APR-1997 (first entry)
DE Pseudomonas sp. SD705 (FERM BP-4772) lipase.
KW Lipase; detergent; food processing; paper; oil; manufacture
KW vector; recombinant production; transformed host; SD705;
KW Pseudomonas; alcaligenes; SD702; Bacillus; NKS-21.
OS Pseudomonas sp.
PN WO9627002-A1.
PD 06-SEP-1996.
PF 23-FEB-1996; J00426.
PR 27-FEB-1995; JP-038527.
PA (SHOW) SHOWA DENKO KK.
PI Ohno K, Sasuga J, Takada H, Yoneda T;
DR WPI; 96-412770/41.
DR N-PSDB; T36695.
PT Gene encoding lipase isolated from Pseudomonas SD705 - useful
PT industrially in detergents, food processing and paper and oil
PT manufacture
PS Claim 3; Pages 32-34; 52pp; Japanese.

```
CC   The present sequence is a lipase isolated from Pseudomonas sp.
CC   SD705 (FERM BP-4772), which is industrially useful in detergents,
CC   food processing, paper manufacture and oil manufacture, etc. . A
CC   vector containing the lipase gene can be used for the recombinant
CC   production of the lipase in transformed hosts, especially
CC   Pseudomonas SD705, P. alcaligenes SD702 or Bacillus NKS-21.
SQ   Sequence   335 AA;


  Query Match             7.2%;  Score 81;  DB 1;  Length 335;
  Best Local Similarity   24.9%;  Pred. No. 1.3;
  Matches   50;  Conservative   32;  Mismatches   85;  Indels   34;  Gaps    8;

Qy      21 RVSFLGAQLPPEVAAMARLLGDLDRSTFRKLLKFVVSSLQGEDCRDGVQRLGVSANLPEE 80
           | |   ||   | ||        :: |  :|      ||  :|  : ::||     ||:|
Db     168 RASLFNAQ---EHAAFFASEEVYNQFTLERLAILHDPSLDPQDKAERIERL--REGLPDE 222

Qy      81 QLGALLAGMH-TLLQQALRLPPTSLKPDTFRDQLQELCIPQDLVGDLASVVFGSQRPLLD 139
             |:  :| || ||  :|    :|    :||::| :  :||| |:        |:
Db     223 LQQLLVPQLHLTLRQQTQQLLEQGAEP----EQLRQLRL--NLVGPQAT-------ERLE 269

Qy     140 SVAQQQGAWLPHVADFRWRVDVAISTSALARSLQPSVLMQLKLSDGSAYRFEVPTAKFQE 199
           :: :|:  |   :: |       ||   || | : : :  |     | :         :
Db     270 ALDRQRSEWDQRLSGFNRERQAIISQPGLADSDKQAAIEALLHEQFSEH---------ER 320

Qy     200 LRYSVALVLKEMADLEKRCER 220
           || |       :  |: | ||
Db     321 LRVS------SLLGLDSRAER 335


RESULT  14
P91395
ID   P91395 standard; peptide; 383 AA.
AC   P91395;
DT   15-FEB-1990  (first entry)
DE   Peptide expressed by E.coli Bio F gene.
KW   E.coli; Bio F; biotin.
OS   Escherichia coli.
PN   GB2216530-A.
PD   11-OCT-1989.
PF   17-MAR-1989; 006210.
PR   22-MAR-1988; GB-006804.
PR   17-MAR-1989; GB-006210.
PA   (UKAG-) UK Min. Agric. Fish.
PI   Pearson BM, McKee RA;
DR   WPI; 89-295085/41.
DR   N-PSDB; N91333.
PT   Plasmid contg. gene(s) for expression of biotin synthetase enzymes
PT   - derived from E.coli and capable of replication and expression in other
PT   microorganisms, esp. yeast.
PS   Table 6; page 39-40; 52pp; English.
CC   The enzyme is expressed by the E.coli bio F gene.  It is a biotin
CC   synthetase enzyme, and permits the transformant to grow in the absence of
CC   exogenous biotin, giving a useful marker for yeast in fermentation.
SQ   Sequence   383 AA;


  Query Match             7.2%;  Score 81;  DB 1;  Length 383;
  Best Local Similarity   21.0%;  Pred. No. 1.5;
  Matches   57;  Conservative   31;  Mismatches   99;  Indels   84;  Gaps   12;

Qy       6 AATPYL---HHP--------------GDSHSGRVSFLGAQLPPEVAAMARLLGDLDRST 47
           ::  ||   |||               |   || ||           :|  ||
Db      45 SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLG------ 99

Qy      48 FRKLLKFVVSSLQGEDCRDGVQRLGVSANLPEEQLGALLAGMHTLLQQALRLPPTSLKPD 107
           : : | |:     |     |    ::| : :|   |     |  | :|  | |: |:
Db      99 YSRALLFI----SGFAANQAV----IAAMMAKEDRIAATGLSHASLLEAASLSPSQLRRF 150

Qy     108 TFRD------QLQELCIPQDLVGDLASVVF---GSQRPL-LDSVAQQQGAWLPHVADFRW 157
             |       |   | | :|  :   ||   |   || :  | ||   ||
Db     151 AHNDVTHLARLLASPCPGQQMV--VTEGVFSMDGDSAPLRIQQVTQQHNGWL-------- 201

Qy     158 RVDVAISTSALA---------RSLQPSVLM----------------------QLKLSDG 185
           || |  |' :           : ::| :|:                        |: :
Db     201 MVDDAHGTGVIGEQGRGSCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARH 260

Qy     186 SAYRFEVPTAKFQELRYSVALVLKEMADLEK 216
             |  :| |: | || |:|:: :  |  :
Db     261 LIYSTSMPPAQAQALRASLAVIRSDEGDARR 291


RESULT  15
R47213
ID   R47213 standard; Protein; 344 AA.
AC   R47213;
DT   27-JUL-1994  (first entry)
DE   Lipase modulator.
KW   Pseudomonas; cloning; lipase gene; lipase modulator gene; lipids;
KW   oils; fats; detergents; diagnostic reagents.
OS   Pseudimonas pseudoalcaligenes.
PN   WO9402617-A.
PD   03-FEB-1994.
PF   23-JUL-1993; E01995.
PR   23-JUL-1992; EP-202281.
PA   (KONN ) GIST-BROCADES NV.
PI   Cox MMJ,  Gerritse G,  Quax WJ;
DR   WPI; 94-048875/06.
DR   N-PSDB; Q56106.
PT   Increasing lipase prodn. in Pseudomonas species - by cloning a
PT   lipase gene and a lipase modulator gene into a homologous
PT   Pseudomonas species
PS   Disclosure; Page 23; 44pp; English.
CC   The lipase gene from Pseudomonas pseudoalcaligenes M1 was cloned in
CC   E. coli as in EP-334462.  The DNA sequence of insert PvuI/EcoRI
CC   contains two open reading frames, one encoding the lipase gene, the
CC   other encoding a putative lipase modulator gene which increases the
CC   lipase productivity in a homologous host cell.  The lipase obtd. is
CC   used for hydrolysing lipids in e.g. fats and oil processing,
CC   detergents or diagnostic reagents.
CC   See also R47212.
SQ   Sequence   344 AA;


  Query Match             7.0%;  Score 79.5;  DB 1;  Length 344;
  Best Local Similarity   28.1%;  Pred. No. 1.9;
  Matches   38;  Conservative   26;  Mismatches   50;  Indels   21;  Gaps    7;

Qy      58 SLQGEDCRDGVQRLGVSANLPEEQLGALLAGMHTLLQQALRLPPTSLKPD-TFRDQLQEL 116
           :|  :   : : ||   |:|||| | ||||     ||  ||    :|:       |:|:|
Db     209 TLDDQQKAEALDRL--RASLPEE-LQALLAPQ---LQAELRQQTAALQAQGASAAQIQQL 262

Qy     117 CIPQDLVGDLASVVFGSQRPLLDSVAQQQGAWLPHVADFRWRVDVAISTSALARSLQPSV 176
           :   ||| |:         |::: ||:  |   :||:|      ::   |: | : :
Db     263 RL--QLVGAEATA-------RLEALDQQRQQWRQRLADYRREKARVLANDGLSESDKQAA 313

Qy     177 LMQLKLSDGSAYRFE 191
           : ::|     :| ||:
Db     314 IAEL-----AAQRFD 323


Search completed: May 15, 2000, 02:06:08
Job time: 10078 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: May 15, 2000, 03:22:30 ; Search time 35.18 Seconds
(without alignments)
92.094 Million cell updates/sec

Title: US-09-223-796-4
Perfect score: 1132
Sequence: 1 MSAVGAATPYLHHPGDSHSG.....ALVLKEMADLEKRCERRLQD 224

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 143561 seqs, 14463640 residues

Total number of hits satisfying chosen parameters: 143561

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	96	8.5	195	2	US-08-822-260-3	Sequence 3, Appli
2	93.5	8.3	195	2	US-08-822-260-1	Sequence 1, Appli
3	82	7.2	363	1	US-08-530-950-6	Sequence 6, Appli
4	82	7.2	399	1	US-08-530-950-10	Sequence 10, Appl
5	82	7.2	399	2	US-08-874-186-92	Sequence 92, Appl
6	81.5	7.2	384	2	US-08-401-068-2	Sequence 2, Appli
7	81.5	7.2	384	2	US-08-846-338-2	Sequence 2, Appli
8	81	7.2	335	2	US-08-875-062-3	Sequence 3, Appli
9	79	7.0	393	1	US-08-530-950-8	Sequence 8, Appli
10	77	6.8	340	1	US-08-606-888A-5	Sequence 5, Appli
11	71.5	6.3	1103	1	US-08-455-543A-53	Sequence 53, Appl
12	71.5	6.3	1103	2	US-08-223-305C-53	Sequence 53, Appl
13	70.5	6.2	489	3	US-08-369-822C-27	Sequence 27, Appl
14	70.5	6.2	3218	1	US-08-764-100-27	Sequence 27, Appl
15	70	6.2	1106	1	US-08-435-675B-5	Sequence 5, Appli
16	70	6.2	1106	1	US-08-336-257A-8	Sequence 8, Appli
17	69.5	6.1	306	1	US-08-328-322-2	Sequence 2, Appli
18	69.5	6.1	915	1	US-08-328-322-5	Sequence 5, Appli
19	69	6.1	395	1	US-08-318-947A-18	Sequence 18, Appl
20	69	6.1	395	2	US-08-795-303-18	Sequence 18, Appl
21	69	6.1	446	3	US-08-956-254-2	Sequence 2, Appli
22	69	6.1	448	2	US-09-015-815-1	Sequence 1, Appli
23	69	6.1	1086	5	5386025-8	Patent No. 5386025
24	69	6.1	2476	2	US-08-276-967-2	Sequence 2, Appli
25	69	6.1	3075	2	US-08-460-309-5	Sequence 5, Appli
26	69	6.1	3075	2	US-08-125-077-5	Sequence 5, Appli
27	67.5	6.0	551	2	US-08-436-771-9	Sequence 9, Appli
28	67.5	6.0	551	2	US-08-434-998-9	Sequence 9, Appli
29	67.5	6.0	551	2	US-08-487-797-9	Sequence 9, Appli
30	67.5	6.0	551	4	PCT-US95-02058-9	Sequence 9, Appli
31	67	5.9	1100	1	US-08-357-598-11	Sequence 11, Appl
32	67	5.9	1100	2	US-09-003-289-11	Sequence 11, Appl
33	67	5.9	1100	4	PCT-US95-16435-11	Sequence 11, Appl
34	66.5	5.9	393	2	US-08-559-303B-74	Sequence 74, Appl
35	66.5	5.9	560	2	US-08-095-728B-6	Sequence 6, Appli
36	66.5	5.9	560	4	PCT-US92-02320A-6	Sequence 6, Appli
37	66.5	5.9	1417	2	US-08-559-303B-78	Sequence 78, Appl
38	66.5	5.9	2629	2	US-08-751-189-4	Sequence 4, Appli
39	66.5	5.9	2629	2	US-09-060-836-4	Sequence 4, Appli
40	66	5.8	15281	2	US-08-471-119A-2	Sequence 2, Appli
41	65.5	5.8	387	5	5240849-5	Patent No. 5240849
42	65.5	5.8	401	2	US-08-591-079-4	Sequence 4, Appli
43	65.5	5.8	494	1	US-08-447-500-2	Sequence 2, Appli
44	65.5	5.8	494	1	US-08-454-097-2	Sequence 2, Appli
45	65.5	5.8	494	1	US-08-447-408-2	Sequence 2, Appli

ALIGNMENTS

```
RESULT   1
US-08-822-260-3
; Sequence 3, Application US/08822260
; Patent No. 5830660
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Goli, Surya K.
;    TITLE OF INVENTION:  NOVEL TUMORIGENESIS PROTEIN
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/822,260
;      FILING DATE:  Herewith
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0247 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  195 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  GenBank
;      CLONE:  265569
US-08-822-260-3


  Query Match             8.5%;  Score 96;  DB 2;  Length 195;
  Best Local Similarity   20.8%;  Pred. No. 0.0017;
  Matches   43;  Conservative   32;  Mismatches   78;  Indels   54;  Gaps    8;
```

```
Qy      41 GDLDRSTFRKLLKFVVSSLQGEDCRDGVQRLGVSANLPEEQLGALLAGMHTLLQQALRLP 100
           |  | : |  ||:    ||           : |   |  |       |  |:|   :
Db      18 GSFDSNAFALLLRAAFQSL------------LDARADEAALD------HPYLKQ---ID 55

Qy     101 PTSLK------------------PDTFRDQLQELCIPQDLVGDLASVVFGSQRPLLDSV 141
           |  ||                    |    |::    :: : :|    : : :  |:::
Db      56 PVVLKHCHAAAATCILEAGKHQVDKSTLSTYLEDCKFDRERI-ELFCTEYQNNKNSLETL 114

Qy     142 AQQQGAWLPHVADFRWRVDVAISTSALARSLQPSVLMQLKL--SDGSAY---RFEVPTAK 196
               |  |||: |  ||::  | |: | :  :|  |: | :  :|  :|    |      :
Db     115 LGSIGRSLPHITDVSWRLEYQIKTNQLHKMYRPGYLVTLNVENNDSQSYPEINFSCNMEQ 174

Qy     197 FQELRYSVALVLKEMADLEKRCERRLQ 223
            |:|      : :: |  |  ||  |
Db     175 LQDL-------VGKLKDASKSLERATQ 194


RESULT   2
US-08-822-260-1
; Sequence 1, Application US/08822260
; Patent No. 5830660
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Goli, Surya K.
;    TITLE OF INVENTION:  NOVEL TUMORIGENESIS PROTEIN
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/822,260
;      FILING DATE:  Herewith
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0247 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  195 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  UTRSNOT02
;      CLONE:  2267574
US-08-822-260-1


  Query Match             8.3%;  Score 93.5;  DB 2;  Length 195;
  Best Local Similarity   22.8%;  Pred. No. 0.0034;
  Matches   33;  Conservative   28;  Mismatches   57;  Indels   27;  Gaps    5;

Qy      85 LLAGMHTLLQQALRLPPTSLKPDTF-RDQLQELCIPQDLVGDLASVVFGSQRPLLDSVAQ 143
           | || |   :  |    | |:   | |:::: |       :   :: ||
Db      71 LEAGKHRADKSTL---STYLEDCKFDRERIELFCTEYQNNKNSLEILLGS---------- 118

Qy     144 QQGAWLPHVADFRWRVDVAISTSALARSLQPSVLMQLKLSDGSAYRFEVPTAKFQELRYS 203
             | |||: |  ||::  | |: | |  :|: |: | : :          :  : |: :|
Db     118 -IGRSLPHITDVSWRLEYQIKTNQLHRMYRPAYLVTLSVQN-------TDSPSYPEISFS 169

Qy     204 VAL-----VLKEMADLEKRCERRLQ 223
            ::     :: :: |  |  ||  |
Db     170 CSMEQLQDLVGKLKDASKSLERATQ 194


RESULT   3
US-08-530-950-6
; Sequence 6, Application US/08530950
; Patent No. 5736381
;  GENERAL INFORMATION:
;    APPLICANT:  Davis, Roger J.
;    APPLICANT:  Raingeaud, Joel
;    APPLICANT:  Gupta, Shashi
;    APPLICANT:  Derijard, Benoit
;    TITLE OF INVENTION:  CYTOKINE-, STRESS-, AND
;    TITLE OF INVENTION:  ONCOPROTEIN-ACTIVATED HUMAN PROTEIN KINASE
;    TITLE OF INVENTION:  KINASES
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/530,950
;      FILING DATE:  19-SEP-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fasse, J. Peter
;      REGISTRATION NUMBER:  32,983
;      REFERENCE/DOCKET NUMBER:  07917/010001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  363 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
US-08-530-950-6


  Query Match             7.2%;  Score 82;  DB 1;  Length 363;
  Best Local Similarity   20.2%;  Pred. No. 0.2;
  Matches   59;  Conservative   44;  Mismatches   89;  Indels  100;  Gaps   13;

Qy       3 AVGAATPYLHHPGDSHSGRVSFLGAQLPPEVAAMARLLGDLDRSTFRKLLKFVVSSLQGE 62
             | |:    :| |    ||           :: |: |:   :|    ::|| :   ::
Db      76 AYGSVNKMVHKP----SG-----------QIMAVKRIRSTVDEKEQKQLLMDLDVVMRSS 120

Qy      63 DC----------RDG-----VQRLGVSAN------------LPEEQLGAL-LAGMHTL 92
           ||          |:|     :: :  | :            :||| || : || :  |
Db     121 DCPYIVQFYGALFREGDCWICMELMSTSFDKFYKYVYSVLDDVIPEEILGKITLATVKAL 180

Qy      93 --LQQALRLPPTSLKPDTFRDQLQELCIPQDLVGDLASVVFGSQRPLLDSVAQQQGAWL- 150
             |:: |::    :||          |  |  |::    ||    |:||:|:: |
Db     181 NHLKENLKIIHRDIKPSN---------ILLDRSGNIKLCDFGISGQLVDSIAKTRDAGCR 231

Qy     150 PHVADFR---------------WRVDVAISTSALARSLQP---SVLMQL---------K 181
           |::|  |               | : : :   |  |   |   ||  ||         :
```

```
Db    232 PYMAPERIDPSASRQGYDVRSDVWSLGITLYELATGRFPYPKWNSVFDQLTQVVKGDPPQ 291

Qy    182 LSDGSAYRF--------------EVPTAKFQELRYSVALVLKEMADLEKRC 218
          ||:     |              |    |::||     ::: |   :|  |
Db    292 LSNSEEREFSPSFINFVNLCLTKDESKRPKYKELLKHPFILMYEERAVEVAC 343


RESULT   4
US-08-530-950-10
; Sequence 10, Application US/08530950
; Patent No. 5736381
;  GENERAL INFORMATION:
;    APPLICANT:  Davis, Roger J.
;    APPLICANT:  Raingeaud, Joel
;    APPLICANT:  Gupta, Shashi
;    APPLICANT:  Derijard, Benoit
;    TITLE OF INVENTION:  CYTOKINE-, STRESS-, AND
;    TITLE OF INVENTION:  ONCOPROTEIN-ACTIVATED HUMAN PROTEIN KINASE
;    TITLE OF INVENTION:  KINASES
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/530,950
;      FILING DATE:  19-SEP-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fasse, J. Peter
;      REGISTRATION NUMBER:  32,983
;      REFERENCE/DOCKET NUMBER:  07917/010001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  399 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
US-08-530-950-10


  Query Match             7.2%;  Score 82;  DB 1;  Length 399;
  Best Local Similarity   20.2%;  Pred. No. 0.23;
  Matches   59;  Conservative   44;  Mismatches   89;  Indels  100;  Gaps   13;

Qy      3 AVGAATPYLHHPGDSHSGRVSFLGAQLPPEVAAMARLLGDLDRSTFRKLLKFVVSSLQGE 62
          | |:    :| |    ||           :: |: |:   :|    ::||  :   ::
Db    112 AYGSVNKMVHKP----SG-----------QIMAVKRIRSTVDEKEQKQLLMDLDVVMRSS 156

Qy     63 DC-----------RDG-----VQRLGVSAN------------LPEEQLGAL-LAGMHTL 92
          ||           |:|     :: :  | :           :||| || : || :  |
Db    157 DCPYIVQFYGALFREGDCWICMELMSTSFDKFYKYVYSVLDDVIPEEILGKITLATVKAL 216

Qy     93 --LQQALRLPPTSLKPDTFRDQLQELCIPQDLVGDLASVVFGSQRPLLDSVAQQQGAWL- 150
            |:: |::    :||           |  |  |::    ||    |:||:|: : |
Db    217 NHLKENLKIIHRDIKPSN---------ILLDRSGNIKLCDFGISGQLVDSIAKTRDAGCR 267

Qy    150 PHVADFR---------------WRVDVAISTSALARSLQP---SVLMQL---------K 181
          |::|  |               | : : :   |  |   |   ||  ||         :
Db    268 PYMAPERIDPSASRQGYDVRSDVWSLGITLYELATGRFPYPKWNSVFDQLTQVVKGDPPQ 327
```

```
Qy    182 LSDGSAYRF--------------EVPTAKFQELRYSVALVLKEMADLEKRC 218
          ||:     |              |    |::||     ::: |   :|  |
Db    328 LSNSEEREFSPSFINFVNLGLTKDESKRPKYKELLKHPFILMYEERAVEVAC 379


RESULT   5
US-08-874-186-92
; Sequence 92, Application US/08874186
; Patent No. 5989885
;  GENERAL INFORMATION:
;    APPLICANT:  Teng, David H-F.
;    APPLICANT:  Tavtigian, Sean V.
;    APPLICANT:  Perry III, William L.
;    APPLICANT:  Skolnick, Mark H.
;    TITLE OF INVENTION:  SPECIFIC MUTATIONS OF MAP KINASE KINASE
;    TITLE OF INVENTION:  4 (MKK4) IN HUMAN TUMOR CELL LINES IDENTIFY IT AS A TUMOR
;    TITLE OF INVENTION:  SUPPRESSOR IN VARIOUS TYPES OF CANCER
;    NUMBER OF SEQUENCES:  96
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Venable, Baetjer, Howard & Civiletti, LLP
;      STREET:  1201 New York Avenue, N.W., Suite 1000
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/874,186
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,482
;      FILING DATE:  10-JAN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Saxe, Stephen A.
;      REGISTRATION NUMBER:  38,609
;      REFERENCE/DOCKET NUMBER:  24884-121392-01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-962-4848
;      TELEFAX:  202-962-8300
;  INFORMATION FOR SEQ ID NO:  92:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  399 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-874-186-92


  Query Match             7.2%;  Score 82;  DB 2;  Length 399;
  Best Local Similarity   20.2%;  Pred. No. 0.23;
  Matches   59;  Conservative   44;  Mismatches   89;  Indels  100;  Gaps   13;

Qy      3 AVGAATPYLHHPGDSHSGRVSFLGAQLPPEVAAMARLLGDLDRSTFRKLLKFVVSSLQGE 62
          | |:    :| |    ||           :: |: |:   :|    ::||  :   ::
Db    112 AYGSVNKMVHKP----SG-----------QIMAVKRIRSTVDEKEQKQLLMDLDVVMRSS 156

Qy     63 DC-----------RDG-----VQRLGVSAN------------LPEEQLGAL-LAGMHTL 92
          ||           |:|     :: :  | :           :||| || : || :  |
Db    157 DCPYIVQFYGALFREGDCWICMELMSTSFDKFYKYVYSVLDDVIPEEILGKITLATVKAL 216

Qy     93 --LQQALRLPPTSLKPDTFRDQLQELCIPQDLVGDLASVVFGSQRPLLDSVAQQQGAWL- 150
            |:: |::    :||           |  |  |::    ||    |:||:|: : |
Db    217 NHLKENLKIIHRDIKPSN---------ILLDRSGNIKLCDFGISGQLVDSIAKTRDAGCR 267

Qy    150 PHVADFR---------------WRVDVAISTSALARSLQP---SVLMQL---------K 181
          |::|  |               | : : :   |  |   |   ||  ||         :
Db    268 PYMAPERIDPSASRQGYDVRSDVWSLGITLYELATGRFPYPKWNSVFDQLTQVVKGDPPQ 327
```

```
Qy    182 LSDGSAYRF--------------EVPTAKFQELRYSVALVLKEMADLEKRC 218
          ||:     |               |    |::||     ::: |   :|  |
Db    328 LSNSEEREFSPSFINFVNLCLTKDESKRPKYKELLKHPFILMYEERAVEVAC 379


RESULT   6
US-08-401-068-2
; Sequence 2, Application US/08401068
; Patent No. 5859335
;  GENERAL INFORMATION:
;    APPLICANT:  Patton, David
;    TITLE OF INVENTION:  Enhanced Biotin Biosynthesis in Plant Tissue
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/401,068
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/351,970
;      FILING DATE:  08-DEC-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  384 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-401-068-2


  Query Match             7.2%;  Score 81.5;  DB 2;  Length 384;
  Best Local Similarity   19.9%;  Pred. No. 0.25;
  Matches   56;  Conservative   32;  Mismatches   90;  Indels  103;  Gaps   12;

Qy      6 AATPYL---HHP--------------GDSHSGRVSFLGAQLPPEVAAMARLLGDLDRST 47
          ::  ||   |||              |   || ||           :|  ||
Db     45 SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLG------ 99

Qy     48 FRKLLKF---------VVSSLQGEDCRDGVQRLGVSANLPEEQLGALLAGMHTLLQQALR 98
          : : | |          |::::  :: |    ||                  |  | :|
Db     99 YSRALLFISGFAANQAVIAAMMAKEDRIAADRLS----------------HASLLEAAS 141

Qy     99 LPPTSLKPDTFRD------QLQELCIPQDLVGDLASVVF---GSQRPL--LDSVAQQQGA 147
          | |: |:     |       |   | | :|  :   ||   |  ||  :  | ||
Db    142 LSPSQLRRFAHNDVTHLARLLASPCPGQQMV--VTEGVFSMDGDSAPLAEIQQVTQQHNG 199

Qy    148 WLPHVADFRWRVDVAISTSALA---------RSLQPSVLM------------------- 179
          ||         || |  |  :           : ::| :|:
Db    200 WL--------MVDDAHGTGVIGEQGRGSCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTV 251

Qy    179 ---QLKLSDGSAYRFEVPTAKFQELRYSVALVLKEMADLEK 216
              |: :    |   :| |: | || |:|::  :  |  :
Db    252 ADYLLQFARHLIYSTSMPPAQAQALRASLAVIRSDEGDARR 292


RESULT   7
```

```
US-08-846-338-2
; Sequence 2, Application US/08846338
; Patent No. 5869719
;  GENERAL INFORMATION:
;    APPLICANT:  Patton, David
;    TITLE OF INVENTION:  Transgenic Plants Having Increased Biotin Content
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5869719artis Corporation
;      STREET:  520 White Plains Road, P.O. Box 2005
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/846,338
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  384 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-846-338-2


  Query Match             7.2%;  Score 81.5;  DB 2;  Length 384;
  Best Local Similarity   19.9%;  Pred. No. 0.25;
  Matches   56;  Conservative   32;  Mismatches   90;  Indels  103;  Gaps   12;

Qy      6 AATPYL---HHP--------------GDSHSGRVSFLGAQLPPEVAAMARLLGDLDRST 47
          ::  ||   |||              |   || ||           :|  ||
Db     45 SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLG------ 99

Qy     48 FRKLLKF---------VVSSLQGEDCRDGVQRLGVSANLPEEQLGALLAGMHTLLQQALR 98
          : : | |          |::::  :: |    ||                  |  | :|
Db     99 YSRALLFISGFAANQAVIAAMMAKEDRIAADRLS----------------HASLLEAAS 141

Qy     99 LPPTSLKPDTFRD------QLQELCIPQDLVGDLASVVF---GSQRPL--LDSVAQQQGA 147
          | |: |:     |       |   | | :|  :   ||   |  ||  :  | ||
Db    142 LSPSQLRRFAHNDVTHLARLLASPCPGQQMV--VTEGVFSMDGDSAPLAEIQQVTQQHNG 199

Qy    148 WLPHVADFRWRVDVAISTSALA---------RSLQPSVLM------------------- 179
          ||         || |  |  :           : ::| :|:
Db    200 WL--------MVDDAHGTGVIGEQGRGSCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTV 251

Qy    179 ---QLKLSDGSAYRFEVPTAKFQELRYSVALVLKEMADLEK 216
              |: :    |   :| |: | || |:|::  :  |  :
Db    252 ADYLLQFARHLIYSTSMPPAQAQALRASLAVIRSDEGDARR 292


RESULT   8
US-08-875-062-3
; Sequence 3, Application US/08875062B
; Patent No. 5942431
; GENERAL INFORMATION:
;  APPLICANT: YONEDA, TADASHI
;  APPLICANT: HARUM, TAKADA
;  APPLICANT: KEI, OHNO
;  APPLICANT: JUNJI, SASUGA
;  TITLE OF INVENTION: NOVEL LIPASE GENE AND PROCESS FOR THE
```

```
;  TITLE OF INVENTION: PRODUCTION OF LIPASE WITH THE USE OF THE SAME
;  FILE REFERENCE: 5059.204
;  CURRENT APPLICATION NUMBER: US/08/875,062B
;  CURRENT FILING DATE: 1997-07-10
;  EARLIER APPLICATION NUMBER: PCT/JP96/00426
;  EARLIER FILING DATE: 1996-02-23
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
;   LENGTH: 335
;   TYPE: PRT
;   ORGANISM: Pseudomonas sp.
US-08-875-062-3


  Query Match             7.2%;  Score 81;  DB 2;  Length 335;
  Best Local Similarity   24.9%;  Pred. No. 0.23;
  Matches   50;  Conservative   32;  Mismatches   85;  Indels   34;  Gaps    8;

Qy      21 RVSFLGAQLPPEVAAMARLLGDLDRSTFRKLLKFVVSSLQGEDCRDGVQRLGVSANLPEE 80
           | |   ||   | ||        :: |  :|      ||  :|  : ::||     ||:|
Db     168 RASLFNAQ---EHAAFFASEEVYNQFTLERLAILHDPSLDPQDKAERIERL--REGLPDE 222

Qy      81 QLGALLAGMH-TLLQQALRLPPTSLKPDTFRDQLQELCIPQDLVGDLASVVFGSQRPLLD 139
              |:  :| || ||  :|     :|    :||::| :  :|||  |:          |:
Db     223 LQQLLVPQLHLTLRQQTQQLLEQGAEP----EQLRQLRL--NLVGPQAT-------ERLE 269

Qy     140 SVAQQQGAWLPHVADFRWRVDVAISTSALARSLQPSVLMQLKLSDGSAYRFEVPTAKFQE 199
           :: :|:  |   :: |       ||   || | : : :  |     | :         :
Db     270 ALDRQRSEWDQRLSGFNRERQAIISQPGLADSDKQAAIEALLHEQFSEH---------ER 320

Qy     200 LRYSVALVLKEMADLEKRCER 220
           || |       :  |: | ||
Db     321 LRVS------SLLGLDSRAER 335


RESULT   9
US-08-530-950-8
; Sequence 8, Application US/08530950
; Patent No. 5736381
;  GENERAL INFORMATION:
;    APPLICANT:  Davis, Roger J.
;    APPLICANT:  Raingeaud, Joel
;    APPLICANT:  Gupta, Shashi
;    APPLICANT:  Derijard, Benoit
;    TITLE OF INVENTION:  CYTOKINE-, STRESS-, AND
;    TITLE OF INVENTION:  ONCOPROTEIN-ACTIVATED HUMAN PROTEIN KINASE
;    TITLE OF INVENTION:  KINASES
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/530,950
;      FILING DATE:  19-SEP-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fasse, J. Peter
;      REGISTRATION NUMBER:  32,983
;      REFERENCE/DOCKET NUMBER:  07917/010001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  393 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
US-08-530-950-8


  Query Match             7.0%;  Score 79;  DB 1;  Length 393;
  Best Local Similarity   20.0%;  Pred. No. 0.5;
  Matches   59;  Conservative   42;  Mismatches   88;  Indels  106;  Gaps   14;

Qy       3 AVGAATPYLHHPGDSHSGRVSFLGAQLPPEVAAMARLLGDLDRSTFRKLLKFVVSSLQGE 62
           | |:    :| |    ||           :: |: |:   :|    ::||  :   ::
Db     106 AYGSVNKMVHKP----SG----------QIMAVKRIRSTVDEKEQKQLLMDLDVVMRSS 150

Qy      63 DC----------RDG-----VQRLGVSAN------------LPEEQLGAL-LAGMHTL 92
           ||          |:|     :: :  | :             :||| || : || :  |
Db     151 DCPYIVQFYGALFREGDCWICMELMSTSFDKFYKYVYSVLDDVIPEEILGKITLATVKAL 210

Qy      93 --LQQALRLPPTSLKPDTF---RDQLQELCIPQDLVGDLASVVFGSQRPLLDSVAQQQGA 147
             |:: |::    :||      |  : :||              ||    |:||:|: : |
Db     211 MHLKENLKIIHRDIKPSNILLDRSGMIKLC-----------DFGISGQLVDSIAKTRDA 258

Qy     148 -----WLPHVADFR-----------WRVDVAISTSALARSLQP---SVLMQL------- 181
                 |   ||             | : : :   |  |   |   ||  ||
Db     259 GCRPYMAPERIDFSASRQGYDVRSDVWSLGITLYELATGRFPYPKWNSVFDQLTQVVKGD 318

Qy     181 --KLSDGSAYRF--------------EVPTAKFQELRYSVALVLKEMADLEKRC 218
             :||:     |               |    |::||     ::: |   :|  |
Db     319 PPQLSNSEEREFSPSFINFVNLCLTKDESKRPKYKELLKHPFILMYEERAVEVAC 373


RESULT  10
US-08-606-888A-5
; Sequence 5, Application US/08606888A
; Patent No. 5766913
;  GENERAL INFORMATION:
;    APPLICANT:  Lin, Shuen-Fuh
;    APPLICANT:  Chiou, Chien-Ming
;    APPLICANT:  Chuang, Kuang-Hsiang
;    TITLE OF INVENTION:  CLONING, EXPRESSION AND NUCLEOTIDE
;    TITLE OF INVENTION:  SEQUENCE OF A AKALINE GENE FROM PSEUDOMONAS
;    TITLE OF INVENTION:  PSEUDOALCALIGENES F-111
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/606,888A
;      FILING DATE:  26-FEB-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Tsao, Rocky Y.
;      REGISTRATION NUMBER:  34,053
;      REFERENCE/DOCKET NUMBER:  06840/003001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;     LENGTH: 340 amino acids
```

```
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-606-888A-5


  Query Match             6.8%;  Score 77;  DB 1;  Length 340;
  Best Local Similarity   26.2%;  Pred. No. 0.69;
  Matches   60;  Conservative   23;  Mismatches   74;  Indels   72;  Gaps   12;

Qy      24 FLGAQLPPEVAAMARLLGDLDRSTFRK--LLKF-----VVSSLQGEDCR-DGVQRLGVSA 75
           :| :||    ||:|: |  |||    |  |::      :|: | |   | | |||| |
Db     121 YLASQL--RDAALAQALDLLDRYIDYKTELVELERRFPMVTELDGLRAREDAVQRLRASL 178

Qy      76 NLPEEQLGALLAGMHTLLQQALR----LPPTSLKPDT-----------FRDQLQELCIPQ 120
             :|   |  |       |  |     |   || | |             |:||:| :||
Db     179 FNAQEH-AAFFASEEVYNQFTLERLAILHDPSLDPQTQAERIERLREGLPDELQQLLVPQ 237

Qy     121 -----------------DLVGDLASVVFGSQR----PLLD-----------------S 140
                           : :  |   :|| |       ||                 :
Db     238 LHLTLRHDPAVADQGAEPEQLRQLRLNLFGPQATERLERLDRQRSEWDQRLERFNRERQA 297

Qy     141 VAQQQGAWLPHVADFRWRVDVAISTSALARSLQPSVLMQLKLSDGSAYR 189
           :  | |||    |  | |:  |      ||:|: : |::|   |  | |
Db     298 IISQPGAW---TATSRPRLRPAARAVQRARALRVNSLLEL---DSRAER 340


RESULT  11
US-08-455-543A-53
; Sequence 53, Application US/08455543A
; Patent No. 5792846
;  GENERAL INFORMATION:
;    APPLICANT:  Harpold, Michael
;    APPLICANT:  Ellis, Steven
;    APPLICANT:  Williams, Mark
;    APPLICANT:  Feldman, Daniel
;    APPLICANT:  McCue, Ann
;    APPLICANT:  Brenner, Robert
;    TITLE OF INVENTION:  HUMAN CALCIUM CHANNEL COMPOSITIONS AND
;    TITLE OF INVENTION:  METHODS
;    NUMBER OF SEQUENCES:  57
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brown, Martin, Haller & McClain
;      STREET:  1660 Union Street
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92101-2926
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/455,543A
;      FILING DATE:  May 31, 1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/223,305
;      FILING DATE:  April 4, 1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/868,354
;      FILING DATE:  April 10, 1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/745,206
;      FILING DATE:  15-AUG-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/620,250
;      FILING DATE:  30-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/482,384
;      FILING DATE:  20-FEB-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/603,751
;      FILING DATE:  04-APR-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US89/01408
;      FILING DATE:  04-APR-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/176,899
;      FILING DATE:  04-APR-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seidman, Stephanie L.
;      REGISTRATION NUMBER:  33,779
;      REFERENCE/DOCKET NUMBER:  6362-52517
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619)238-0999
;      TELEFAX:  (619)238-0062
;  INFORMATION FOR SEQ ID NO:  53:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1103 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FRAGMENT TYPE:  internal
US-08-455-543A-53


  Query Match             6.3%;  Score 71.5;  DB 1;  Length 1103;
  Best Local Similarity   23.1%;  Pred. No. 18;
  Matches   31;  Conservative   19;  Mismatches   59;  Indels   25;  Gaps    5;

Qy       9 PYLHHPGDSHSGRVSFLGAQLPPEVAAMAR---LLGDLDRSTFRKLLK---------- 54
             | : :|       : || |:|  ::    |   : |:    ||| |:|
Db     546 PNIQNPKSQEPVTLDFLDAELENDIKVEIRNKMIDGESGEKTFRTLVKSQDERYIDKGNR 605

Qy      54 -FVVSSLQGEDCRDGVQRLGVSANLPEEQLGALLAGMHTLLQQALRLPPTSLKPDTFRDQ 112
            :  : : | |          ::  ||      : | :   : || |   | |||| | :
Db     606 TYTWTPVNGTD-------YSLALVLPTYSFYYIKAKLEETITQA-RYSET-LKPDNFEES 656

Qy     113 LQELCIPQDLVGDL 126
                 |:|   ||
Db     657 GYTFIAPRDYCNDL 670


RESULT  12
US-08-223-305C-53
; Sequence 53, Application US/08223305C
; Patent No. 5851824
;  GENERAL INFORMATION:
;    APPLICANT:  Harpold, Michael
;    APPLICANT:  Ellis, Steven
;    APPLICANT:  Williams, Mark
;    APPLICANT:  Feldman, Daniel
;    APPLICANT:  McCue, Ann
;    APPLICANT:  Brenner, Robert
;    TITLE OF INVENTION:  HUMAN CALCIUM CHANNEL COMPOSITIONS AND
;    TITLE OF INVENTION:  METHODS
;    NUMBER OF SEQUENCES:  57
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brown, Martin, Haller & McClain
;      STREET:  1660 Union Street
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92101-2926
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/223,305C
;      FILING DATE:  April 4, 1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/868,354
```

```
;       FILING DATE:  April 10, 1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/745,206
;       FILING DATE:  15-AUG-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/620,250
;       FILING DATE:  30-NOV-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/482,384
;       FILING DATE:  20-FEB-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/603,751
;       FILING DATE:  04-APR-1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  WO PCT/US89/01408
;       FILING DATE:  04-APR-1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/176,899
;       FILING DATE:  04-APR-1988
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Seidman, Stephanie L.
;       REGISTRATION NUMBER:  33,779
;       REFERENCE/DOCKET NUMBER:  52516  (P519739)
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (619)238-0999
;       TELEFAX:  (619)238-0062
;   INFORMATION FOR SEQ ID NO:  53:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1103 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
;     FRAGMENT TYPE:  internal
US-08-223-305C-53


  Query Match             6.3%;  Score 71.5;  DB 2;  Length 1103;
  Best Local Similarity   23.1%;  Pred. No. 18;
  Matches   31;  Conservative   19;  Mismatches   59;  Indels   25;  Gaps    5;

Qy       9 PYLHHPGDSHSGRVSFLGAQLPPEVAAMAR---LLGDLDRSTFRKLLK------------ 54
           | : :|       : || |:|  ::    |   : |:    ||| |:|
Db     546 PNIQNPKSQEPVTLDFLDAELENDIKVEIRNKMIDGESGEKTFRTLVKSQDERYIDKGNR 605

Qy      54 -FVVSSLQGEDCRDGVQRLGVSANLPEEQLGALLAGMHTLLQQALRLPPTSLKPDTFRDQ 112
            :  : : | |         ::  ||      : | :   : || |   | |||| | :
Db     606 TYTWTPVNGTD-------YSLALVLPTYSFYYIKAKLEETITQA-RYSET-LKPDNFEES 656

Qy     113 LQELCIPQDLVGDL 126
                |:|   ||
Db     657 GYTFIAPRDYCNDL 670


RESULT  13
US-08-369-822C-27
; Sequence 27, Application US/08369822C
; Patent No. 6015660
;  GENERAL INFORMATION:
;    APPLICANT:  Lipkin, W. I.
;    APPLICANT:  Briese, Thomas
;    APPLICANT:  Kliche, Stefanie
;    APPLICANT:  Schneider, Patrick A.
;    APPLICANT:  Stitz, Lothar
;    APPLICANT:  Schneemann, Anette
;    TITLE OF INVENTION:  Borna Disease Viral Sequences,
;    TITLE OF INVENTION:  Diagnostics and Therapeutics for Central Nervous
;    TITLE OF INVENTION:  System Diseases
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fulbright & Jaworski, L.L.P.
;      STREET:  865 South Figueroa Street, 29th Floor
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  90017-2571
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  WINDOWS NT-WORDPERFECT 8.0
;      SOFTWARE:  ASCII (DOS) TEXT
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/369,822C
;      FILING DATE:  06-JAN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Churchill, Margaret A. (Ph.D.)
;      REGISTRATION NUMBER:  39,944
;      REFERENCE/DOCKET NUMBER:  1279-194XX
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  213/892-9200
;      TELEFAX:  213/680-4518
;  INFORMATION FOR SEQ ID NO:  27:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  489 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-369-822C-27


  Query Match             6.2%;  Score 70.5;  DB 3;  Length 489;
  Best Local Similarity   26.0%;  Pred. No. 6.9;
  Matches   34;  Conservative   15;  Mismatches   49;  Indels   33;  Gaps    6;

Qy      13 HPGDSHSGRVSFL-GAQLPPEVAAMARLLGDLDRSTFRKLLKFVVSSLQGEDCRDGVQRL 71
           | |  : |:  :| | |||  :  |||  | |  | |             :| :  :  :
Db     349 HSGFIYFGKKQYLNGVQLPQSLKTMAR-CGPLSDSIF------------DDLQGSLASI 394

Qy      72 GVSANLPEEQLGAL-----LAGMHTLL------QQALRLP-------PTSLKPDTFRDQL 113
           | |       :   :     :|  |::|      |  | |           || || ::|
Db     395 GTSFERGTSETRHIFPSRWIASFHSMLAINLLNQNHLGFPLGFNIDISCFKKPLTFSEKL 454

Qy     114 QELCIPQDLVG 124
             |  || | |
Db     455 IALITPQVLGG 465


RESULT  14
US-08-764-100-27
; Sequence 27, Application US/08764100
; Patent No. 5773700
;  GENERAL INFORMATION:
;    APPLICANT:  van Grinsven J., Martinus Q.
;    APPLICANT:  De Haan, Petrus T.
;    APPLICANT:  Gielen L., Johannes J.
;    APPLICANT:  Peters, Dirk
;    APPLICANT:  Goldbach, Robert W.
;    TITLE OF INVENTION:  Improvements in or Relating to Organic
;    TITLE OF INVENTION:  Compounds
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sandoz Agro, Inc
;      STREET:  975 California Avenue
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/08/764,100
;      FILING DATE:  06-DEC-1996
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/214,064
;      FILING DATE:
;      APPLICATION NUMBER:  US 08/032,235
;      FILING DATE:  17-MAR-1993
;      APPLICATION NUMBER:  GB 9206016.9
;      FILING DATE:  19-MAR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5773700ris, Allen E.
;      REGISTRATION NUMBER:  34,490
;      REFERENCE/DOCKET NUMBER:  137-1061
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 354-3592
;      TELEFAX:  (415) 857-1125
;  INFORMATION FOR SEQ ID NO:  27:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3218 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-764-100-27


  Query Match             6.2%;  Score 70.5;  DB 1;  Length 3218;
  Best Local Similarity   22.1%;  Pred. No. 1.2e+02;
  Matches   58;  Conservative   41;  Mismatches   94;  Indels   69;  Gaps   13;

Qy       7 ATPYLHHPGDSHSGRVSFLGAQLPPEVA----------AMA-RLLGDLDRSTFRKLLKFV 55
             |:|         | ||: ||  |  |||          |:|  ||  | |     :|: :
Db    2839 ASPLYSSERTYRSERVA-LGLALALEVALASPSERVALALAVALLEGLPRGLASNILEIL 2897

Qy      56 VSS--LQGEDCRDGVQRLGVSANLPEEQLGALLAGMHTLLQQALRLPPTSLKPDTFRDQL 113
            :|  |   :   |:  :   |::    ||  | |::|      |   :     | | :
Db    2898 EASPGLNASNASPGLNHISALAHI--SGLGLGLNGLNT-----YRASNSERASPTHRSER 2950

Qy     114 QELCIPQDLVGDL--------ASVVFGSQRP----------LLDSVAQQ------QGAWL 149
             |  | :   :          :::: |:||          :|:::| :       :| :
Db    2951 VALTRPSERPHETRPASPTYRVALLYSSERPRPHEASNPHEILEALASERHISPHEGLYS 3010

Qy     150 -----PHVADFRWRVDVAISTSALARSLQPSVLMQLKLSDGSAYRFEVPTAKFQELRYSV 204
                 || |    ||       |||| :    |::|:: :   :   :|
Db    3011 ERPHEPHEASPTHRV--------ALARGVALVALLELEILELEPHEVALP---------HE 3054

Qy     205 ALVLKEMADLEKRC---ERRLQ 223
           || |: :|    |   || |:
Db    3055 ALALEALATYRLECYSSERILE 3076


RESULT  15
US-08-435-675B-5
; Sequence 5, Application US/08435675B
; Patent No. 5710250
;  GENERAL INFORMATION:
;    APPLICANT:  Ellis, Steven Bradley
;    APPLICANT:  Williams, Mark E.
;    APPLICANT:  Harpold, Michael Miller
;    APPLICANT:  Schwartz, Arnold
;    APPLICANT:  Brenner, Robert
;    TITLE OF INVENTION:  CALCIUM CHANNEL COMPOSITIONS AND METHODS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brown, Martin, Haller & McClain
;      STREET:  1660 Union Street
;      CITY:  San Diego
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92101-2926
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/435,675B
;      FILING DATE:  05-MAY-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/314,083
;      FILING DATE:  28-SEP-1994
;      APPLICATION NUMBER:  US 07/914,231
;      FILING DATE:  13-JUL-1992
;      APPLICATION NUMBER:  US 07/603,751
;      FILING DATE:  08-NOV-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seidman, Stephanie L
;      REGISTRATION NUMBER:  33,779
;      REFERENCE/DOCKET NUMBER:  6362-53193
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-238-0999
;      TELEFAX:  619-238-0062
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1106 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FRAGMENT TYPE:  internal
US-08-435-675B-5


  Query Match             6.2%;  Score 70;  DB 1;  Length 1106;
  Best Local Similarity   23.7%;  Pred. No. 28;
  Matches   32;  Conservative   18;  Mismatches   59;  Indels   26;  Gaps    5;

Qy       9 PYLHHPGDSHSGRVSFLGAQLPPEVAAMAR---LLGDLDRSTFRKLLKFVVSSLQGEDCR 65
           | : :|       : || |:|  ::    |   : |:    ||| |:|       :| |
Db     548 PNVQNPKSQEPVTLDFLDAELENDIKVEIRNKMIDGESGEKTFRTLVK------SQDER 600

Qy      66 -------------DGVQRLGVSANLPEEQLGALLAGMHTLLQQALRLPPTSLKPDTFRD 111
                        :|   :  ::  ||      : | :   : || |   | |||| | :
Db     601 YIDKGNRTYTWTPVNGTDYSSLALVLPTYSFYYIKAKIEETITQA-RYSET-LKPDNFEE 658

Qy     112 QLQELCIPQDLVGDL 126
                 |:|  ||
Db     659 SGYTFLAPRDYCSDL 673


Search completed: May 15, 2000, 03:22:32
Job time: 5778 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: May 15, 2000, 03:23:35 ; Search time 52.16 Seconds
(without alignments)
251.784 Million cell updates/sec

Title: US-09-223-796-4
Perfect score: 1132
Sequence: 1 MSAVGAATPYLHHPGDSHSG.....ALVLKEMADLEKRCERRLQD 224

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 168808 seqs, 58629743 residues

Total number of hits satisfying chosen parameters: 168808

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : PIR_63:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	95	8.4	200	2	S72612	RNA helicase II -
2	93.5	8.3	676	2	S41022	hypothetical prote
3	86.5	7.6	332	2	T00657	hypothetical prote
4	86	7.6	863	2	T38016	importin beta-1 su
5	85.5	7.6	2182	2	T14320	calcineurin inhibi
6	83	7.3	218	2	T35174	hypothetical prote
7	82	7.2	399	2	I38901	JNK-activating pro
8	82	7.2	997	2	A60776	230k bullous pemph
9	81.5	7.2	384	1	SYECKP	8-amino-7-oxononan
10	81.5	7.2	800	2	S13032	3',5'-cyclic-GMP p
11	81.5	7.2	856	2	S30762	3',5'-cyclic-GMP p
12	81.5	7.2	1337	2	A53824	nuclear pore membr
13	81	7.2	145	2	E69440	conserved hypothet
14	81	7.2	544	2	S61303	heat shock protein
15	80.5	7.1	161	2	E72339	hypothetical prote
16	80.5	7.1	446	2	S34997	pfeS protein - Pse
17	80	7.1	278	2	B25977	beta-glucoside pos
18	80	7.1	311	3	T31438	exopolyphosphatase
19	80	7.1	1947	1	S05697	myosin heavy chain
20	80	7.1	1968	2	T22232	hypothetical prote
21	79.5	7.0	829	2	T40239	probable helicase
22	79.5	7.0	890	2	G64740	uridylyltransferas
23	79.5	7.0	918	2	S04255	regulatory protein
24	79.5	7.0	1321	2	T10929	3C3.20c protein -
25	79.5	7.0	1386	2	T00257	hypothetical prote
26	79	7.0	853	2	A36617	3',5'-cyclic-GMP p
27	78.5	6.9	653	2	A60961	leishmanolysin (EC
28	78	6.9	234	2	S35020	nolW protein - Rhi
29	78	6.9	415	2	C75592	conserved hypothet
30	78	6.9	458	2	A75386	conserved hypothet
31	78	6.9	541	2	S61301	heat shock protein
32	78	6.9	1241	2	S40528	phosphorylase kina
33	78	6.9	4151	2	T13734	groovin gene prote
34	77.5	6.8	377	2	C65051	rubredoxin--NAD+ r
35	77.5	6.8	690	2	D75487	v-type ATP synthas
36	77.5	6.8	971	2	T33907	hypothetical prote
37	77.5	6.8	1218	2	T29915	hypothetical prote
38	77.5	6.8	3329	2	T30904	breast cancer tumo
39	77	6.8	195	1	IVHO22	interferon alpha-I
40	77	6.8	230	2	D64080	uroporphyrin-III C
41	77	6.8	318	2	E70609	probable folP2 pro
42	77	6.8	373	2	S48228	3-isopropylmalate
43	77	6.8	375	2	T03593	leucoanthocyanidin
44	77	6.8	544	2	S61300	heat shock protein
45	77	6.8	666	2	T35864	hypothetical prote

ALIGNMENTS

RESULT 1
S72612
RNA helicase II - human (fragments)
C;Species: Homo sapiens (man)
C;Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 05-Dec-1998
C;Accession: S72612
R;Valdez, B.C.; Henning, D.; Busch, R.K.; Woods, K.; Flores-Rozas, H.; Hurwitz, J.; P
Nucleic Acids Res. 24, 1220-1224, 1996
A;Title: A nucleolar RNA helicase recognized by autoimmune antibodies from a patient
A;Reference number: PC6010; MUID:96188853
A;Accession: S72612
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-16;17-31;32-47;48-55;56-66;67-87;88-99;100-105;106-119;120-144;145-163;

Query Match 8.4%; Score 95; DB 2; Length 200;
Best Local Similarity 28.8%; Pred. No. 0.36;
Matches 51; Conservative 24; Mismatches 58; Indels 44; Gaps 10;

```
Qy      72 GVSANLP--EE--QLGALLAGMHT------LLQQALRLPPTSLKPDT-FRDQLQELCIPQ 120
           | |:| |  ||  :|   |  : |      |:::|: |       |  | ||: |    |
Db      33 GASSNFPISEETIKLVTFLFPIQTFSFAIPLIEKAVVLDEVDQMLDMGFADQVSETTYEQ 92

Qy     121 -DLVGDLASVVFGSQRPLLDS-VAQQQGAWLPHVADFRWRVDVAISTS------------ 167
            ||:|  |:|:  :|    :| : |  |  :|        ||: | :|
Db      93 VDLIGKKAAVIADAQELSQNSAIKQDAGLDIP-------EVDLVIQSSPPKDVESYIHRG 145

Qy     167 ---ALARSL-----QPSVLMQLKLSDGSAYRFEVPTAKFQELRYSVALVLKEMADLE 215
               ||| :|      ||:: |||       |:|||  |::     |  |  :||
Db     146 AVEALAAALAHISGATSVMVFLKL----GVCFDVPTASVTEIQEKQLSVATEQPELE 198
```

RESULT 2
S41022
hypothetical protein T07C4.9 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Sep-1997
C;Accession: S41022
R;Berks, M.
submitted to the EMBL Data Library, January 1994
A;Reference number: S41014
A;Accession: S41022
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-676 <BER>
A;Cross-references: EMBL:Z29443; NID:g1067051; PID:g443836
C;Genetics:
A;Introns: 69/3; 161/1; 208/3; 227/1; 357/1; 505/3
C;Superfamily: annexin repeat homology
F;373-444/Domain: annexin repeat homology <AX1>
F;445-516/Domain: annexin repeat homology <AX6>
F;528-600/Domain: annexin repeat homology <AX3>

F;604-675/Domain: annexin repeat homology <AX4>

Query Match 8.3%; Score 93.5; DB 2; Length 676;
Best Local Similarity 22.7%; Pred. No. 2.2;
Matches 41; Conservative 32; Mismatches 63; Indels 45; Gaps 8;

```
Qy      43 LDRSTFRKLLKFVVSSLQGEDCRDGVQRLGVSANLPEEQLGALLAGMHTLLQQALRLPPT 102
           |: | |||| : |  |   |            .  | |||    || ::     : ||
Db      50 LNNSKFRKLCELVKESRSHE------------MLSEEQ----LATLYDCATTSTRL--- 90

Qy     103 SLKPDTFRDQLQELCIPQDLV--GDLASVVFG---------------SQRPLLDSVAQQQ 145
             |: | |  ::|   ::::  |:|                      ::  |: : :|:
Db      90 ---PNDFADDQEDL---KNIIRYGELFKACHAINSTDFIQKSEDLKDEEKVALERIVEQK 143

Qy     146 GA-WLPHVADFRWRVDVAISTSALARSLQPSVLMQLKLSDGSAYRFEVPTAKFQELRYSV 204
            |    :  |  | |:: :: | :|:||:| : :|:    |    ||     | : |  :
Db     144 LAESAKNEKDIAWNVNIVVANSYVAKSLRPVINIQMPTVGGDT-NFEFDIDSFAQFRQQL 202

Qy     205 A 205
           |
Db     203 A 203
```

RESULT 3
T00657
hypothetical protein F3I6.20 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 22-Oct-1999
C;Accession: T00657
R;Federspiel, N.A.; Palm, C.J.; Conway, A.B.; Kurtz, D.B.; Conway, A.R.; Au, M.; Araujo,
; Vysotskaia, V.S.; Yu, G.; Ecker, J.; Theologis, A.; Davis, R.W.
submitted to the EMBL Data Library, February 1998
A;Reference number: Z14197
A;Accession: T00657
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-332 <FED>
A;Cross-references: EMBL:AC002396; NID:g2749918; PID:g2829879; GSPDB:GN00059; ATSP:F3I6.
C;Genetics:
A;Gene: ATSP:F3I6.20
A;Map position: 1
A;Introns: 15/1; 34/3; 60/3; 86/1; 107/3; 137/3; 162/3; 173/3; 194/3; 217/3; 232/3; 265/

Query Match 7.6%; Score 86.5; DB 2; Length 332;
Best Local Similarity 22.2%; Pred. No. 3.7;
Matches 51; Conservative 27; Mismatches 75; Indels 77; Gaps 8;

```
Qy      28 QLPPEVAAMARLLG---DLDRSTFRKLLKF---------------VVSSLQGEDCR-DG 67
           :|  |:  | | |    :: : | ::: |                  :|  : || :|
Db     112 ELSSEIDMMGRRLDANTEVIQETIQEVAKLQDGTSFIKDDVKAVFDAFENLASKVCRIEG 171

Qy      68 VQRLGVSANLPEEQLGALLAGMHTLLQQALRLPPTSLKPDTFRDQLQELCIPQDLVGDLA 127
            | |  :::||     || |       : | ||| |  ||
Db     172 NQALPSTSSLP-----ALEAAPMAPSSKTLSLPPAS--PDE------------------- 206

Qy     128 SVVFGSQRPLLDSVAQQQGAWLPHVADFRWRVDVAISTSAL-------------ARSLQP 174
                || |   :|||:    | |        |:  |:|:                |
Db     206 -----SQSPSTPNVAQKSRGLLQHTQSMSGLKDINESSSSHNTSSNGIYFGGNGASGSSS 260

Qy     175 SVLMQLKLSDGSAYRFEVPTAKFQELRYSVALVLKEMADLEKRCERRLQD 224
            || ::    |  |  |              ::|||: | ||| :  ||:
Db     261 GVLGRINKEGGMLYLVEY------------VLLKELLDNEKRLKVNLQE 297
```

RESULT 4
T38016
importin beta-1 subunit - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T38016
R;Oliver, K.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, August 1995
A;Reference number: Z21762
A;Accession: T38016
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-863 <OLI>
A;Cross-references: EMBL:Z98532; PIDN:CAB11082.1; GSPDB:GN00066; SPDB:SPAC1B1.03c
A;Experimental source: strain 972h-; cosmid c1B1
C;Genetics:
A;Gene: SPDB:SPAC1B1.03c
A;Map position: 1

Query Match 7.6%; Score 86; DB 2; Length 863;
Best Local Similarity 20.9%; Pred. No. 13;
Matches 42; Conservative 36; Mismatches 73; Indels 50; Gaps 6;

```
Qy      33 VAAMARLLGDLDRSTFRKLLKFVVSSLQGEDCR-----DGVQRLGVSANLPEEQLGALLA 87
           : :: |  |   |::  :::  :: ::|    :     | :  :|   |  |||    :
Db     580 LTSIIRRFGPDIRTSSDQIMNLLLQTMQTAPKQSVVHEDVLLAIGAMMNSLEEQFEVYVP 639

Qy      88 GMHTLLQQALRLPPTSLKPDTFRDQLQELC-IPQDLVGDLASVVFGSQRPLLDSVAQQQG 146
             |  ||             :|  :|| :   ||||||  :
Db     640 SFVPFLSSALS----------NEQEYQLCSVAVGLVGDLARAL-------------NA 674

Qy     147 AWLPHVADFRWRVDVAISTSALARSLQPSVL---MQLKLSDGSAYRFEV---------- 193
            ||:  ||  |:   : :| | |:::|::|     : |: |:|::  :
Db     675 KILPYCDDFMTRLVQDLQSSVLDRNVKPAILSCFSDIALAIGAAFQTYLEAVMVLLQQAS 734

Qy     193 -----PTAKFQELRYSVALVL 208
                | | |  : |  || |
Db     735 SVQAPPGANFSMIDYVDALRL 755
```

RESULT 5
T14320
calcineurin inhibitor cain - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T14320
R;Lai, M.M.; Burnett, P.E.; Wolosker, H.; Blackshaw, S.; Snyder, S.H.
J. Biol. Chem. 273, 18325-18331, 1998
A;Title: Cain, a novel physiologic protein inhibitor of calcineurin.
A;Reference number: Z17979; MUID:98325042
A;Accession: T14320
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-2182 <LAI>
A;Cross-references: EMBL:AF061947; NID:g3323606; PID:g3323607; PIDN:AAC40176.1

Query Match 7.6%; Score 85.5; DB 2; Length 2182;
Best Local Similarity 20.8%; Pred. No. 47;
Matches 49; Conservative 34; Mismatches 90; Indels 63; Gaps 8;

```
Qy      10 YLHHPGDSHSGRVSFLGAQLPPEV-----------AAMARLLGDLDRSTFRK-LLKFVV 56
           |||      :  :: :      |||:           |:: :|||  |     : |: |:
Db    1236 YLHEEAARYPKKIHY---HNPPELAMEALEVYFRLHASILKLLGKPDSGVSAEVLVSFMK 1292

Qy      57 SSLQG----------------------EDCRDGVQRL-GVSANLPEEQLGALLAGMHTL 92
            : :|                      ||     | |  |:||     | :  :|
Db    1293 EAAEGPFARGEEKNTPKASEKEKACLVDEDSHSSAGTLPGPGASLPSSSGPGLTSPPYTA 1352

Qy      93 L----------QQALRLPPTSLKPDTFRDQLQELCIPQDLVGDLASVVFGSQRPL-- 138
                       |||    |     |:    | :    ||   :  |:: ||:: |
Db    1353 TPIDHDYVKCKKPRQQA---TPDDRSQDSTAVALSDSSSTQDFFNEPTSLLDGSRKLLPE 1409

Qy     138 -----LDSVAQQQGAWLPHVADFRWRVDVAISTSALARSLQPSVLMQLKLSDGSAY 188
                | : |   |  ||   : | : : :: ::   | ::|||   :  |  |||
Db    1410 KRISGLSAQAGPSGKDLPGPTEERGKTEESLESTEAFRVVEPSVQKPVADSSASAY 1465
```

RESULT 6

T35174
hypothetical protein SC5A7.25c SC5A7.25c - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C;Accession: T35174
R;Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, July 1998
A;Reference number: Z21570
A;Accession: T35174
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-218 <SEE>
A;Cross-references: EMBL:AL031107; PIDN:CAA19954.1; GSPDB:GN00070; SCOEDB:SC5A7.25c
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SC5A7.25c

Query Match 7.3%; Score 83; DB 2; Length 218;
Best Local Similarity 27.1%; Pred. No. 4.4;
Matches 48; Conservative 22; Mismatches 51; Indels 56; Gaps 9;

```
Qy      68 VQRLGVSANLPEEQLGALLAGMHT--------LLQQALRLPPTSLKPDTFRDQLQELCIP 119
Db      40 VGRLSAASGVPESVVGALLGGRPTGGADVQTRFVQRLDLLRRTRLKPNGRKYTQQEIA-- 98

Qy     120 QDLVG----DLASVVFGSQRPLLDSVAQQQGAWLPHVADFRWRVDVAISTSALARSLQPS 175
Db      98 -DGAGMSRQQAGALINGDRRPTMEHCDAIQRFFRVH-AGFL----TAEDPEALAGALQ-- 150

Qy     176 VLMQLKLSDGSAYRFEVPTAKFQELRYSVALVLKEMADLEKRC--------ERRLQD 224
Db     150 ------------RTE------QEL-------LQQLADRERQAAAPADDPLERLLQD 180
```

RESULT 7
I38901
JNK-activating protein kinase (EC 2.7.1.-) - human
N;Alternate names: dual specificity kinase JNKK; MAP kinase kinase MKK4
C;Species: Homo sapiens (man)
C;Date: 16-Feb-1996 #sequence_revision 16-Feb-1996 #text_change 18-Jun-1999
C;Accession: I38901; B55556
R;Lin, A.; Minden, A.; Martinetto, H.; Claret, F.X.; Lange-Carter, C.; Mercurio, F.; Joh
Science 268, 286-290, 1995
A;Title: Identification of a dual specificity kinase that activates the Jun kinases and
A;Reference number: A56160; MUID:95232504
A;Accession: I38901
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-399 <RES>
A;Cross-references: EMBL:U17743; NID:g791187; PIDN:AAC50127.1; PID:g791188
A;Note: it is uncertain whether Met-1 or Met-37 is the initiator
R;Derijard, B.; Raingeaud, J.; Barrett, T.; Wu, I.H.; Han, J.; Ulevitch, R.J.; Davis, R.
Science 267, 682-685, 1995
A;Title: Independent human MAP kinase signal transduction pathways defined by MEK and MK
A;Reference number: A55556; MUID:95141073
A;Accession: B55556
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 37-399 <EER>
A;Cross-references: GB:L36870
C;Genetics:
A;Gene: JNKK1
C;Superfamily: kinase-related transforming protein; protein kinase homology
C;Keywords: ATP; phosphotransferase
F;100-367/Domain: protein kinase homology <KIN>
F;108-116/Region: protein kinase ATP-binding motif

Query Match 7.2%; Score 82; DB 2; Length 399;
Best Local Similarity 20.2%; Pred. No. 11;
Matches 59; Conservative 44; Mismatches 89; Indels 100; Gaps 13;

```
Qy       3 AVGAATPYLHHPGDSHSGRVSFLGAQLPPEVAAMARLLGDLDRSTFRKLLKFVVSSLQGE 62
Db     112 AYGSVNKMVHKP----SG-----------QIMAVKRIRSTVDEKEQKQLLMDLDVVMRSS 156

Qy      63 DC-----------RDG-----VQRLGVSAN-------------LPEEQLGAL-LAGMHTL 92
Db     157 DCPYIVQFYGALFREGDCWICMELMSTSFDKFYKYVYSVLDDVIPEEILGKITLATVKAL 216

Qy      93 --LQQALRLPPTSLKPDTFRDQLQELCIPQDLVGDLASVVFGSQRPLLDSVAQQQGAWL- 150
Db     217 NHLKENLKIIHRDIKPSN---------ILLDRSGNIKLCDFGISGQLVDSIAKTRDAGCR 267

Qy     150 PHVADFR----------------WRVDVAISTSALARSLQP---SVLMQL---------K 181
Db     268 PYMAPERIDPSASRQGYDVRSDVWSLGITLYELATGRFPYPKWNSVFDQLTQVVKGDPPQ 327

Qy     182 LSDGSAYRF---------------EVPTAKFQELRYSVALVLKEMADLEKRC 218
Db     328 LSNSEEREFSPSFINFVNLCLTKDESKRPKYKELLKHPFILMYEERAVEVAC 379
```

RESULT 8
A60776
230k bullous pemphigoid antigen BPM1 - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 28-May-1993 #sequence_revision 28-May-1993 #text_change 28-May-1993
C;Accession: A60776
R;Amagai, M.; Hashimoto, T.; Tajima, S.; Inokuchi, Y.; Shimizu, N.; Saito, M.; Miki,
J. Invest. Dermatol. 95, 252-259, 1990
A;Title: Partial cDNA cloning of the 230-kD mouse bullous pemphigoid antigen by use o
A;Reference number: A60776; MUID:90347244
A;Accession: A60776
A;Molecule type: mRNA
A;Residues: 1-997 <AMA>
C;Keywords: glycoprotein; membrane protein

Query Match 7.2%; Score 82; DB 2; Length 997;
Best Local Similarity 22.4%; Pred. No. 35;
Matches 59; Conservative 30; Mismatches 106; Indels 68; Gaps 11;

```
Qy       2 SAVGAATPYLHHPGDSHSGRVSFL-------------------------GAQLPPEVAA 35
Db     228 SAFDTAGRECHHPAEISPGNSGHLNLKTRLPLSRWTQEPHQTEGKWPHRAAEQLPKEV-- 286

Qy      36 MARLLG-DLDRSTFRKLLKFVVSSLQGE-----DCRDGVQRLGVSANLPEEQLGALLAGM 89
Db     286 QFRQPGAPLDRESSQPCYSEDFSQTSTELQITFDDKNPITR--------------LSEL 330

Qy      90 HTLLQQALR--LPPTSLKPDTFRDQLQELCIPQDLV-----GDLASVVFGSQRPLLDSVA 142
Db     331 ETMREQALHPSRPPVTYQDDKLERELVKLLTPLEIAKNKQCGMHTEVTTLKQEKRLGSSA 390

Qy     143 QQQGAWLPHVADFRWRVDVAISTSALARSLQPSVLMQLKLSDGSAYRFEVPTAKFQELRY 202
Db     391 ---GGWMLEGC----RTSGGLKGDFLKKSVEPEASPSLDLNQACSVRDE--EFQFQGLRH 441

Qy     203 SVA---LVLKEMADLEKRCERRL 222
Db     442 TVTGRQLVEAKLLDMRTVEQLRL 464
```

RESULT 9
SYECKP
8-amino-7-oxononanoate synthase (EC 2.3.1.47) - Escherichia coli
N;Alternate names: 7-KAP synthetase; 7-keto-8-amino pelargonic acid synthetase
C;Species: Escherichia coli
C;Date: 30-Jun-1990 #sequence_revision 30-Jun-1990 #text_change 18-Jun-1999
C;Accession: D32025; H64813
R;Otsuka, A.J.; Buoncristiani, M.R.; Howard, P.K.; Flamm, J.; Johnson, C.; Yamamoto,
J. Biol. Chem. 263, 19577-19585, 1988
A;Title: The Escherichia coli biotin biosynthetic enzyme sequences predicted from the
A;Reference number: A32025; MUID:89066784

```
A;Accession: D32025
A;Molecule type: DNA
A;Residues: 1-384 <OTS>
A;Cross-references: GB:J04423; NID:g145422; PIDN:AAA23516.1; PID:g145426
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617
A;Accession: H64813
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-384 <BLAT>
A;Cross-references: GB:AE000180; GB:U00096; NID:g1786988; PIDN:AAC73863.1; PID:g1786993;
A;Experimental source: strain K-12, substrain MG1655
C;Genetics:
A;Gene: bioF
A;Map position: 17 min
C;Superfamily: 5-aminolevulinate synthase; glycine C-acetyltransferase homology
C;Keywords: acyltransferase; biotin biosynthesis; phosphoprotein; pyridoxal phosphate
F;44-378/Domain: glycine C-acetyltransferase homology <GCA>
F;236/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted


  Query Match             7.2%;  Score 81.5;  DB 1;  Length 384;
  Best Local Similarity   19.9%;  Pred. No. 12;
  Matches   56;  Conservative   32;  Mismatches   90;  Indels  103;  Gaps   12;

Qy       6 AATPYL---HHP---------------GDSHSGRVSFLGAQLPPEVAAMARLLGDLDRST 47
           ::  ||   |||               |   || ||           :|  ||
Db      45 SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLG------ 99

Qy      48 FRKLLKF---------VVSSLQGEDCRDGVQRLGVSANLPEEQLGALLAGMHTLLQQALR 98
           : : | |         |::::  :: |    ||                  |  | :|
Db      99 YSRALLFISGFAANQAVIAAMMAKEDRIAADRLS----------------HASLLEAAS 141

Qy      99 LPPTSLKPDTFRD------QLQELCIPQDLVGDLASVVF---GSQRPL--LDSVAQQQGA 147
           | |: |:     |       |   |  | :|  :   ||   |   ||  :  | ||
Db     142 LSPSQLRRFAHNDVTHLARLLASPCPGQQMV--VTEGVFSMDGDSAPLAEIQQVTQQHNG 199

Qy     148 WLPHVADFRWRVDVAISTSALA---------RSLQPSVLM------------------- 179
           ||         || |  |  :          : ::| :|:
Db     200 WL--------MVDDAHGTGVIGEQGRGSCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTV 251

Qy     179 ---QLKLSDGSAYRFEVPTAKFQELRYSVALVLKEMADLEK 216
              |: :   |   :| |: | || |:|::  :  |  :
Db     252 ADYLLQFARHLIYSTSMPPAQAQALRASLAVIRSDEGDARR 292


RESULT  10
S13032
3',5'-cyclic-GMP phosphodiesterase (EC 3.1.4.35) beta' chain - mouse
C;Species: Mus musculus (house mouse)
C;Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 20-Aug-1999
C;Accession: S13032; S30763
R;Baehr, W.; Champagne, M.S.; Lee, A.K.; Pittler, S.J.
FEBS Lett. 278, 107-114, 1991
A;Title: Complete cDNA sequences of mouse rod photoreceptor cGMP phosphodiesterase alpha
cing of the beta-subunit gene.
A;Reference number: S13030; MUID:91130581
A;Accession: S13032
A;Molecule type: DNA
A;Residues: 1-800 <BAE>
A;Cross-references: EMBL:X60133
R;Baehr, W.
submitted to the EMBL Data Library, June 1991
A;Reference number: S30762
A;Accession: S30763
A;Molecule type: DNA
A;Residues: 1-559,'R',561-800 <BA2>
A;Cross-references: GB:X87952; EMBL:X60133; NID:g871432; PIDN:CAA61202.1; PID:g871433
C;Superfamily: 3',5'-cyclic-GMP phosphodiesterase alpha chain; 3',5'-cyclic-nucleotide p
C;Keywords: alternative splicing; cGMP binding; phosphoric diester hydrolase


  Query Match             7.2%;  Score 81.5;  DB 2;  Length 800;
  Best Local Similarity   19.4%;  Pred. No. 30;
  Matches   49;  Conservative   38;  Mismatches   68;  Indels   97;  Gaps   10;

Qy      32 EVAAMARLLGDLDRS-----TFRKLLKFVVSSLQGEDC-------RDGVQRLGVSANLPE 79
           | ||:  |:. |:  |         |:|: : : |  : |       |:|:  |
Db      55 ESAALFELVQDMQESVNMERVVFKILRRLCTILHADRCSLFMYRQRNGIAEL-------- 107

Qy      80 EQLGALLAGMHTLLQQALRLPPTSLKPDTFRDQLQELCIPQDLVGDLASVVFGSQRPLLD 139
                        | ||   |::||:    |::  :| |     : :||    ::
Db     107 ----------------ATRL--FSVQPDSL---LEDCLVPPD-----SEIVFPLDIGIVG 140

Qy     140 SVAQ-------QQGAWLPHVADFRWRVDVAISTSALARSLQ-----PSVLMQLKLSDGS- 187
           |||        |  |  || : |    :   :: : |:  :        :|:| :   ||
Db     141 HVAQTKKMINVQDVAECPHFSSFADELTDYVTKNILSTPIMNGKDVVAVIMAVNKLDGPC 200

Qy     187 ------------------------AYRFEVPTAKFQELRYSVALVLKEMADLEKR---- 218
                                    :|    | : | | :|   | :|: |:|::
Db     201 FTSEDEDVFTKYLNFATLNLKIYHLSYLHNCETRRGQVLLWSANKVFEELTDIERQFHKA 260

Qy     218 ---------CER 220
                    |||
Db     261 FYTVRAYLNCER 272


RESULT  11
S30762
3',5'-cyclic-GMP phosphodiesterase (EC 3.1.4.35) beta chain - mouse
C;Species: Mus musculus (house mouse)
C;Date: 22-Nov-1993 #sequence_revision 19-Oct-1995 #text_change 20-Aug-1999
C;Accession: S30762; S13031; S13121
R;Baehr, W.
submitted to the EMBL Data Library, June 1991
A;Reference number: S30762
A;Accession: S30762
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-856 <BAE>
A;Cross-references: EMBL:X60133; NID:g53595; PIDN:CAA42719.1; PID:g53596
R;Baehr, W.; Champagne, M.S.; Lee, A.K.; Pittler, S.J.
FEBS Lett. 278, 107-114, 1991
A;Title: Complete cDNA sequences of mouse rod photoreceptor cGMP phosphodiesterase al
cing of the beta-subunit gene.
A;Reference number: S13030; MUID:91130581
A;Accession: S13031
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-559,'G',561-856 <BA2>
A;Cross-references: EMBL:X60133
R;Bowes, C.; Li, T.; Danciger, M.; Baxter, L.C.; Applebury, M.L.; Farber, D.B.
Nature 347, 677-680, 1990
A;Title: Retinal degeneration in the rd mouse is caused by a defect in the beta subun
A;Reference number: S13121; MUID:91015387
A;Accession: S13121
A;Molecule type: mRNA
A;Residues: 1-2,'X',4,'G',6-18,'S',20-48,'DV',51-157,'T',159-175,'C',177-231,'R',233-
A;Cross-references: EMBL:X55968; NID:g53616; PIDN:CAA39439.1; PID:g53617
A;Note: the authors translated the codon AGA for residue 232 as Glu
C;Superfamily: 3',5'-cyclic-GMP phosphodiesterase alpha chain; 3',5'-cyclic-nucleotid
C;Keywords: cGMP binding; phosphoric diester hydrolase


  Query Match             7.2%;  Score 81.5;  DB 2;  Length 856;
  Best Local Similarity   19.4%;  Pred. No. 32;
  Matches   49;  Conservative   38;  Mismatches   68;  Indels   97;  Gaps   10;

Qy      32 EVAAMARLLGDLDRS-----TFRKLLKFVVSSLQGEDC-------RDGVQRLGVSANLPE 79
           | ||:  |: |:  |         |:|: : : |  : |       |:|:  |
Db      55 ESAALFELVQDMQESVNMERVVFKILRRLCTILHADRCSLFMYRQRNGIAEL-------- 107

Qy      80 EQLGALLAGMHTLLQQALRLPPTSLKPDTFRDQLQELCIPQDLVGDLASVVFGSQRPLLD 139
```

```
                        | ||   |::||:    |::  :| |     : :||     ::
Db    107 ---------------ATRL--FSVQPDSL---LEDCLVPPD-----SEIVFPLDIGIVG 140

Qy    140 SVAQ-------QQGAWLPHVADFRWRVDVAISTSALARSLQ-----PSVLMQLKLSDGS- 187
          |||       |  |  || : |   :   :: : |:  :       :|:| :   ||
Db    141 HVAQTKKMINVQDVAECPHFSSFADELTDYVTKNILSTPIMNGKDVVAVIMAVNKLDGPC 200

Qy    187 ------------------------AYRFEVPTAKFQELRYSVALVLKEMADLEKR---- 218
                                   :|     | : | | :|   | :|: |:|::
Db    201 FTSEDEDVFTKYLNFATLNLKIYHLSYLHNCETRRGQVLLWSANKVFEELTDIERQFHKA 260

Qy    218 ---------CER 220
                   |||
Db    261 FYTVRAYLNCER 272
```

RESULT 12
A53824
nuclear pore membrane protein POM152 - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YM9553.05; protein YMR129w
C;Species: Saccharomyces cerevisiae
C;Date: 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 24-Nov-1999
C;Accession: A53824; B53824; S53059
R;Wozniak, R.W.; Blobel, G.; Rout, M.P.
J. Cell Biol. 125, 31-42, 1994
A;Title: POM152 is an integral protein of the pore membrane domain of the yeast nuclear
A;Reference number: A53824; MUID:94186543
A;Accession: A53824
A;Molecule type: DNA
A;Residues: 1-1337 <WOZ>
A;Cross-references: GB:Z31592; NID:g473153; PID:g473154
A;Accession: B53824
A;Molecule type: protein
A;Residues: 217-237;270-302;330-353;417-434;505-528;546-563;579-595;684-697;944-958;992-
R;Badcock, K.; Churcher, C.
submitted to the EMBL Data Library, March 1995
A;Reference number: S53055
A;Accession: S53059
A;Molecule type: DNA
A;Residues: 1-1337 <BAD>
A;Cross-references: EMBL:Z48622; NID:g728663; PID:g728668; MIPS:YMR129w
C;Genetics:
A;Gene: SGD:POM152
A;Cross-references: SGD:S0004736; MIPS:YMR129w
A;Map position: 13R
C;Superfamily: Saccharomyces cerevisiae nuclear pore membrane protein POM152
C;Keywords: blocked amino end; duplication; glycoprotein; nuclear membrane; transmembran
F;176-195/Domain: transmembrane #status predicted <TMM>
F;280/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match   7.2%; Score 81.5; DB 2; Length 1337;
  Best Local Similarity   25.9%; Pred. No. 56;
  Matches   45; Conservative   28; Mismatches   72; Indels   29; Gaps   9;

```
Qy     56 VSSLQGEDCRDGVQRLGVSANLPEEQLGA---LLAGMHTLLQQALRLPPTSLKPDTFRDQ 112
          ::|: |    |  : :| | |:  |  |     ||    :  |      :||:|: |
Db    380 IASITGTGSND--RCIGDSDNVSFEIQGVPPMKLAYSKIVNGQTFSYVDSSLQPEYFESP 437

Qy    113 LQELCIPQDLV-GDLASVVFGSQRPL---LDSVAQQQGAWLPHVADFRWRVDVAISTSAL 168
          ||     |   |:|  : :|  :|:   |||   | |       | :::|   |  |
Db    438 LQSSKSKQSFTQGELNDLKWGRNQPVNINLDSSITQDG-------KFAYKIDKI--TDGL 488

Qy    169 ARSLQ-PSVLMQLKLSDGSAYRF---EVPTAKFQE-------LRYSVALVLKEM 211
             :   |:  :||     :| |   ||| |  :|        : |:|:| :|:
Db    489 GNVVDFTSLPEELKKRYDLSYNFNVHEVPRAALEERFDPKSPTKRSIAIVFEEI 542
```

RESULT 13
E69440
conserved hypothetical protein AF1526 - Archaeoglobus fulgidus
C;Species: Archaeoglobus fulgidus
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 29-Sep-1999
C;Accession: E69440
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes,
Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch
A;Reference number: A69250; MUID:98049343
A;Accession: E69440
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-145 <KLE>
A;Cross-references: GB:AE000997; GB:AE000782; NID:g2689320; PIDN:AAB89722.1; PID:g264
C;Superfamily: hypothetical protein AF1526

  Query Match   7.2%; Score 81; DB 2; Length 145;
  Best Local Similarity   30.7%; Pred. No. 3.9;
  Matches   35; Conservative   21; Mismatches   40; Indels   18; Gaps   5;

```
Qy     36 MARLLGDLDRS-----TFRKLLKFVVSSLQGEDCRDGVQRLGVSANLPEEQLGALLAGMH 90
          | |:|  || :      |:|| |     |:||     :: : |   :|||:    || |
Db      1 MERILLVLDDTGRGEIAFQKLKKLAEDGLRGEVYILYIREMEVPPFVPEEK---ELAAYH 57

Qy     91 TLLQQALRLPPTSLKPDTFRDQLQELCIPQDLVGDLASVVFGSQRPLLDSVAQQ 144
            |: |:::      | : |::||::  :    | |: |||||     |  | :|
Db     58 RLMTQSMK------KLEGFKNQLEKAGLK---VSDV-SVVFGKYADRLLLVEKQ 101
```

RESULT 14
S61303
heat shock protein 63 - Neisseria meningitidis (fragment)
C;Species: Neisseria meningitidis
C;Date: 27-Apr-1996 #sequence_revision 17-May-1996 #text_change 26-Aug-1999
C;Accession: S61303; S40249
R;Pannekoek, Y.; Dankert, J.; van Putten, J.P.M.
Mol. Microbiol. 15, 277-285, 1995
A;Title: Construction of recombinant neisserial Hsp60 proteins and mapping of antigen
A;Reference number: S61300; MUID:95264914
A;Accession: S61303
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-544 <PAN>
A;Cross-references: EMBL:Z22956; NID:g438205; PIDN:CAA80532.1; PID:g438206
C;Superfamily: chaperonin groEL
C;Keywords: heat shock; stress-induced protein

  Query Match   7.2%; Score 81; DB 2; Length 544;
  Best Local Similarity   19.9%; Pred. No. 20;
  Matches   61; Conservative   45; Mismatches   84; Indels 116; Gaps   14;

```
Qy     25 LGAQLPPEVAAMAR-LLGD-------LDRSTFRKLLKFVVSSLQGEDCRDGVQRL----- 72
          :|||:  |||:    : ||       | :|   : :|:| : :   | : |: :
Db     69 MGAQMVKEVASKTNDVAGDGTTTATVLAQSIVAEGMKYVTAGMNPTDLKRGIDKAVAALV 128

Qy     72 --------------------GVSANLPEEQLGALLA---------GMHTL---------- 93
                             :|||  :||:||::|          |: |:
Db    129 EELKNIAKPCDTSKEIAQVGSISAN-SDEQVGAIIAEAMEKVGKEGVITVEDGKSLENEL 187

Qy     93 -LQQALRLPPTSLKP-------------------------DTFRDQLQE---------- 116
           : : ::     | |                         || |
Db    188 DVVEGMQFDRGYLSPYFINDAEKQIAGLDNPFVLLFEKKISNIRDLLPVLEKVAKASRPL 247

Qy    116 LCIPQDLVGD-LASVVFGSQRPLLDSVAQQQGAWLPHVADFRWRV--DVAISTSALARSL 172
          | | ::: |: ||::|  : | :| :|| :     |   | |  :  |:|| |    |
Db    248 LIIAENVEGEALATLVVNNIRGILKTVAVK----APGFGDRRKAMLQDIAILTGGTVISE 303

Qy    173 QPSVLMQLKLSD--GSAYRFEVP---------------TAKFQELRYSVALVLKEMADL 214
          :  : ::    |  | | | |:                  |:| |:|  :     :  |
Db    304 EVGLFLEKATLDNLGQAKRIEIGKENTTIIDGFGDAAQIEARFAEIRQQIETATSDY-DK 362
```

```
Qy     215 EKRCER 220
           ||  ||
Db     363 EKLQER 368


RESULT  15
E72339
hypothetical protein - Thermotoga maritima (strain MSB8)
C;Species: Thermotoga maritima
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C;Accession: E72339
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
 C.M.
Nature 399, 323-329, 1999
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A;Reference number: A72200; MUID:99287316
A;Accession: E72339
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-161 <ARN>
A;Cross-references: GB:AE001744; GB:AE000512; NID:g4981254; PID:g4981263; TIGR:TM0738
A;Experimental source: strain MSB8
C;Genetics:
A;Gene: TM0738


  Query Match             7.1%;  Score 80.5;  DB 2;  Length 161;
  Best Local Similarity   27.7%;  Pred. No. 5;
  Matches   26;  Conservative   20;  Mismatches   35;  Indels   13;  Gaps    3;

Qy       1 MSAVGAATP-YLHHPGDSHSGRVSFLGAQLPPEVAAMARLLGDLDRSTFRKLLKFVVSSL 59
           ::||   |  ::: ||   | |   ||::|  :   : | : :|:|      ||     |
Db      69 LAAVAGRTQVFIYTPGTKSSERTHSLGSELQDKQQEIERKILELERQ-----LKTETDPL 123

Qy      60 QGEDCRDGVQRLGVSANL-------PEEQLGALL 86
           : |  |: ::|| :: |:       ||  :| ||
Db     124 ERERLREDLERLKLALNILKASLRAPELLVGVLL 157


Search completed: May 15, 2000, 03:23:38
Job time: 5548 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: May 15, 2000, 05:34:20 ; Search time 44.1 Seconds
(without alignments)
154.692 Million cell updates/sec

Title: US-09-223-796-4
Perfect score: 1132
Sequence: 1 MSAVGAATPYLHHPGDSHSG.....ALVLKEMADLEKRCERRLQD 224

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 83857 seqs, 30454973 residues

Total number of hits satisfying chosen parameters: 83857

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : SwissProt_38:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	86	7.6	863	1	IMB1_SCHPO	O13864 schizosacch
2	85.5	7.6	757	1	YHGF_NEIME	Q51152 neisseria m
3	85.5	7.6	2182	1	CABI_RAT	O88480 rattus norv
4	83	7.3	397	1	MPK4_MOUSE	P47809 m dual spec
5	82.5	7.3	964	1	UL70_MCMVS	Q69153 murine cyto
6	82	7.2	399	1	MPK4_HUMAN	P45985 homo sapien
7	81.5	7.2	384	1	BIOF_ECOLI	P12998 escherichia
8	81.5	7.2	1337	1	P152_YEAST	P39685 saccharomyc
9	81	7.2	544	1	CH60_NEIME	P42385 neisseria m
10	80.5	7.1	446	1	PFES_PSEAE	Q04804 pseudomonas
11	80	7.1	278	1	BGLG_ECOLI	P11989 escherichia
12	80	7.1	1947	1	MYSC_CAEEL	P12845 caenorhabdi
13	79.5	7.0	890	1	GLND_ECOLI	P27249 escherichia
14	79.5	7.0	918	1	QA1S_NEUCR	P11637 neurospora
15	79	7.0	853	1	CNRB_BOVIN	P23439 bos taurus
16	78	6.9	234	1	NOLW_RHIFR	P33212 rhizobium f
17	78	6.9	234	1	NOLW_RHISN	P55712 rhizobium s
18	78	6.9	1241	1	KPB1_MOUSE	P18826 mus musculu
19	77.5	6.8	3329	1	BRC2_MOUSE	P97929 mus musculu
20	77	6.8	195	1	INO2_HORSE	P05002 equus cabal
21	77	6.8	230	1	HEMX_HAEIN	P44773 haemophilus
22	77	6.8	318	1	DHP2_MYCTU	O05308 mycobacteri
23	77	6.8	373	1	LEU3_CANMA	P07139 candida mal
24	77	6.8	544	1	CH60_NEIGO	P29842 neisseria g
25	76.5	6.8	302	1	VPJ_BPP2	P51767 bacteriopha
26	76.5	6.8	361	1	RF1_COXBU	P47849 coxiella bu
27	76.5	6.8	377	1	YGBD_ECOLI	P37596 escherichia
28	76.5	6.8	496	1	NIFE_RHISN	P55673 rhizobium s
29	76.5	6.8	3137	1	CA36_CHICK	P15989 gallus gall
30	76.5	6.8	4568	1	DYHB_CHLRE	Q39565 chlamydomon
31	76	6.7	1239	1	V120_EBV	P03189 epstein-bar
32	75.5	6.7	544	1	CH60_NEIFL	P48215 neisseria f
33	75.5	6.7	548	1	CH60_ACYPS	P25750 acyrthosiph
34	75.5	6.7	592	1	LAM2_MOUSE	P21619 mus musculu
35	75.5	6.7	1391	1	YLD5_CAEEL	Q03570 caenorhabdi
36	75	6.6	359	1	COBT_ECOLI	P36562 escherichia
37	75	6.6	493	1	CSD_RAT	Q64611 rattus norv
38	75	6.6	772	1	NFL1_HUMAN	Q14494 h nuclear f
39	75	6.6	1101	1	GUNC_CELFI	P14090 cellulomona
40	75	6.6	2594	1	7LES_DROVI	P20806 drosophila
41	74.5	6.6	1189	1	HAIR_HUMAN	O43593 homo sapien
42	74	6.5	220	1	RP29_HUMAN	O95707 homo sapien
43	74	6.5	446	1	MPK2_XENLA	Q07192 xenopus lae
44	74	6.5	551	1	CH60_BUCAP	Q59177 buchnera ap
45	74	6.5	1301	1	DDX9_CAEEL	Q22307 caenorhabdi

ALIGNMENTS

```
RESULT   1
IMB1_SCHPO
ID   IMB1_SCHPO      STANDARD;      PRT;   863 AA.
AC   O13864;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   IMPORTIN BETA-1 SUBUNIT (KARYOPHERIN BETA-1 SUBUNIT) (IMPORTIN 95).
GN   SPAC1B1.03C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
OC   Schizosaccharomycetaceae; Schizosaccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   Oliver K., Harris D., Barrell B.G., Rajandream M.A., Wood V.;
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: REQUIRED FOR NUCLEAR PROTEIN IMPORT AND MEDIATES DOCKING
CC       OF IMPORT SUBSTRATE TO DISTINCT NUCLEOPORINS. SERVES A RECEPTOR
CC       FOR NUCLEAR LOCALIZATION SIGNALS.
CC   -!- SUBUNIT: FORMS A COMPLEX WITH IMPORTIN ALPHA SUBUNIT (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR ENVELOPE (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE IMPORTIN BETA CHAIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 11 HEAT REPEATS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; Z98532; CAB11082.1; -.
KW   Transport; Protein transport; Repeat; Nuclear protein.
FT   DOMAIN       124    482       HEAT REPEATS DOMAIN 1.
FT   DOMAIN       329    342       IAB DOMAIN.
FT   DOMAIN       334    419       RAN-GTP BINDING (BY SIMILARITY).
FT   DOMAIN       592    717       HEAT REPEATS DOMAIN 2.
FT   DOMAIN       315    320       POLY-ALA.
SQ   SEQUENCE   863 AA;  94747 MW;  9E02D72F5A1C6839 CRC64;

  Query Match          7.6%;  Score 86;  DB 1;  Length 863;
  Best Local Similarity  20.9%;  Pred. No. 4.1;
  Matches   42;  Conservative   36;  Mismatches   73;  Indels   50;  Gaps    6;

Qy      33 VAAMARLLGDLDRSTFRKLLAFVVSSLQGEDCR-----DGVQRLGVSANLPEEQLGALLA 87
           : :: |  |   |::  :::  :: ::|     :     |  :  :|   |  |||    :
Db     580 LTSIIRRFGPDIRTSSDQIMNLLLQTMQTAPKQSVVHEDVLLAIGAMMNSLEEQFEVYVP 639

Qy      88 GMHTLLQQALRLPPTSLKPDTFRDQLQELC-IPQDLVGDLASVVFGSQRPLLDSVAQQQG 146
           |  ||                :|  :|| :   ||||||  :
Db     640 SFVPFLSSALS----------NEQEYQLCSVAVGLVGDLARAL----------NA 674
```

```
Qy      147 AWLPHVADFRWRVDVAISTSALARSLQPSVL---MQLKLSDGSAYRFEV---------- 193
            ..      ||:  ||  |:   : :| | |:::|::|     : |: |:|::  :
Db      675 KILPYCDDFMTRLVQDLQSSVLDRNVKPAILSCFSDIALAIGAAFQTYLEAVMVLLQQAS 734

Qy      193 -----PTAKFQELRYSVALVL 208
                 . | | |  : |  || |
Db      735 SVQAPPGANFSMIDYVDALRL 755


RESULT   2
YHGF_NEIME
ID   YHGF_NEIME     STANDARD;      PRT;   757 AA.
AC   Q51152;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HYPOTHETICAL 83.1 KD PROTEIN IN REGION E.
OS   Neisseria meningitidis.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B1940;
RX   MEDLINE; 96236055.
RA   Petering H., Hammerschmidt S., Frosch M., van Putten J.P.M.,
RA   Ison C.A., Robertson B.D.;
RT   "Genes associated with meningococcal capsule complex are also found
RT   in Neisseria gonorrhoeae.";
RL   J. Bacteriol. 178:3342-3345(1996).
CC   -!- SIMILARITY: CONTAINS A COPY OF THE 'S1 MOTIF'.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L09189; AAC37046.1; -.
DR   HSSP; P05055; 1SRO.
DR   PFAM; PF00575; S1; 1.
KW   Hypothetical protein; RNA-binding.
FT   DOMAIN       640    709       S1 MOTIF.
SQ   SEQUENCE   757 AA;  83064 MW;  EAAF4D9A8197C944 CRC64;

  Query Match             7.6%;  Score 85.5;  DB 1;  Length 757;
  Best Local Similarity  23.2%;  Pred. No. 3.9;
  Matches   49;  Conservative   27;  Mismatches   68;  Indels   67;  Gaps    8;

Qy       28 QLPPEVAAMARLLGDLDRSTFRKLLKFVVSSLQGEDCRDGVQRLGVSANLPEEQLGALLA 87
            :: |:   : :   |::::  | |  ||     |||   |  :||  |     | | ::
Db      448 KIDPKSIGVGQYQHDVNQNQLAKSLDAVV-----EDC---VNAVGVDVNTASAPLLARIS 499

Qy       88 GMHTLLQQ----------------ALRLPPTSLKPDTFRDQLQELCI-----PQDLVGD 125
            |::  | |                |::|   |   ||     | |     | |
Db      500 GLNQTLAQNIVAYRDENGAFDSRKKLLKVP--RLGEKTFEQAAGFLRINGGKEPLD---- 554

Qy      126 LASVVFGSQRPLLDSVAQQQGAWLPHVADFRWRVDVAISTSALARSLQPSVLMQLKLSDG 185
             || |     |::  :  |||     :   | ||                      |:| ||
Db      554 -ASAVHPEAYPVVAKMLAQQGISAAELIGNRERV----------------KQIKASDF 594

Qy      186 SAYRFEVPTAKFQELRYSVALVLKEMADLEK 216
            :  || :||             :|  :::|||
Db      595 TDERFGLPT------------ILDILSELEK 613


RESULT   3
CABI_RAT
ID   CABI_RAT       STANDARD;      PRT;  2182 AA.
AC   O88480;
DT   15-FEB-2000 (Rel. 39, Created)
DT   15-FEB-2000 (Rel. 39, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   CALCINEURIN-BINDING PROTEIN CABIN 1 (CALCINEURIN INHIBITOR) (CAIN).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98325042.
RA   Lai M.M., Burnett P.E., Wolosker H., Blackshaw S., Snyder S.H.;
RT   "Cain, a novel physiologic protein inhibitor of calcineurin.";
RL   J. Biol. Chem. 273:18325-18331(1998).
CC   -!- FUNCTION: INTERACTS WITH AND INHIBITS CALCINEURIN-MEDIATED SIGNAL
CC       TRANSDUCTION.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF061947; AAC40176.1; -.
SQ   SEQUENCE   2182 AA;  242811 MW;  2904C335AB3F440A CRC64;

  Query Match             7.6%;  Score 85.5;  DB 1;  Length 2182;
  Best Local Similarity  20.8%;  Pred. No. 14;
  Matches   49;  Conservative   34;  Mismatches   90;  Indels   63;  Gaps    8;

Qy       10 YLHHPGDSHSGRVSFLGAQLPPEV----------AAMARLLGDLDRSTFRK-LLKFVV 56
            |||     :  :: :    |||:                |:: :|||  |     : |: |:
Db     1236 YLHEEAARYPKKIHY---HNPPELAMEALEVYFRLHASILKLLGKPDSGVSAEVLVSFMK 1292

Qy       57 SSLQG---------------------EDCRDGVQRL-GVSANLPEEQLGALLAGMHTL 92
            : :|                     ||      | |  |:||     : | : :|
Db     1293 EAAEGPFARGEEKNTPKASEKEKACLVDEDSHSSAGTLPGPGASLPSSSGPGLTSPPYTA 1352

Qy       93 L----------QQALRLPPTSLKPDTFRDQLQELCIPQDLVGDLASVVFGSQRPL-- 138
            |          |||    |    |:      | :    ||   :  |:: ||:: |
Db     1353 TPIDHDYVKCKKPRQQA---TPDDRSQDSTAVALSDSSSTQDFFNEPTSLLDGSRKLLPE 1409

Qy      138 -----LDSVAQQQGAWLPHVADFRWRVDVAISTSALARSLQPSVLMQLKLSDGSAY 188
                 | :|   |  ||   : | : : :: ::   | ::|||   :  |  |||
Db     1410 KRISGLSAQAGPSGKDLPGPTEERGKTEESLESTEAFRVVEPSVQKPVADSSASAY 1465


RESULT   4
MPK4_MOUSE
ID   MPK4_MOUSE     STANDARD;      PRT;   397 AA.
AC   P47809;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   DUAL SPECIFICITY MITOGEN-ACTIVATED PROTEIN KINASE KINASE 4
DE   (EC 2.7.1.-) (MAP KINASE KINASE 4) (MAPKK 4) (MAPK/ERK KINASE 4) (JNK
DE   ACTIVATING KINASE 1) (C-JUN N-TERMINAL KINASE KINASE 1) (JNK KINASE 1)
DE   (JNKK 1) (SAPK/ERK KINASE 1) (SEK1).
GN   MAP2K4 OR PRKMK4 OR JNKK1 OR SEK1 OR SERK1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE; 95089821.
RA   Sanchez I., Hughes R.T., Mayer B.J., Yee K., Woodgett J.R., Avruch J.,
RA   Kyriakis J.M., Zon L.I.;
RT   "Role of SAPK/ERK kinase-1 in the stress-activated pathway regulating
RT   transcription factor c-Jun.";
RL   Nature 372:794-798(1994).
RN   [2]
RP   REVISIONS.
```

```
RA   Zon L.I.;
RL   Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: DUAL SPECIFICITY KINASE THAT ACTIVATES THE JUN KINASES
CC       MAPK8 (JNK1) AND MAPK9 (JNK2) AS WELL AS MAPK14 (P38) BUT NOT
CC       MAPK1 (ERK2) OR MAPK3 (ERK1).
CC   -!- PTM: ACTIVATED BY PHOSHORYLATION ON SER/THR BY MAP KINASE KINASE
CC       KINASES (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: IS EXPRESSED ABUNDANTLY IN THE ADULT
CC       BRAIN AND MUSCLE.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC       MAP KINASE KINASE SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U18310; AAB81554.1; -.
DR   HSSP; P00518; 1PHK.
DR   MGD; MGI:107199; SERK1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PFAM; PF00069; pkinase; 1.
KW   Transferase; Serine/threonine-protein kinase; Tyrosine-protein kinase;
KW   ATP-binding; Phosphorylation.
FT   DOMAIN        5     17       GLY/SER-RICH.
FT   DOMAIN      100    366       PROTEIN KINASE.
FT   NP_BIND     103    111       ATP (BY SIMILARITY).
FT   BINDING     129    129       ATP (BY SIMILARITY).
FT   ACT_SITE    227    227       BY SIMILARITY.
FT   MUTAGEN     129    129       K->R: LOSS OF ATP-BINDING ACTIVITY.
SQ   SEQUENCE   397 AA;  44113 MW;  B99C6688184E5B3D CRC64;

  Query Match                    7.3%;  Score 83;  DB 1;  Length 397;
  Best Local Similarity   20.2%;  Pred. No. 2.9;
  Matches   59;  Conservative   44;  Mismatches   89;  Indels  100;  Gaps   13;

Qy       3 AVGAATPYLHHPGDSHSGRVSFLGAQLPPEVAAMARLLGDLDRSTFRKLLKFVVSSLQGE 62
           | |:     :| |    ||                :: |: |:    :|    ::|| :    ::
Db     110 AYGSVNKMVHKP----SG----------          QIMAVKRIRSTVDEKEQKQLLMDLDVVMRSS 154

Qy      63 DC----------RDG-----VQRLGVSAN------------LPEEQLGAL-LAGMHTL 92
           ||          |:|     ::  :  | :                :||| || : || :  |
Db     155 DCPYIVQFYGALFREGDCWICMELMSTSFDKFYKYVYSVLDDVIPEEILGKITLATVKAL 214

Qy      93 --LQQALRLPPTSLKPDTFRDQLQELCIPQDLVGDLASVVFGSQRPLLDSVAQQQGAWL- 150
             |:: |::   :||              |  |  |::    ||     |:||:|: : |
Db     215 NHLKENLKIIHRDIKPSN---------ILLDRSGNIKLCDFGISGQLVDSIAKTRDAGCR 265

Qy     150 PHVADFR----------------WRVDVAISTSALARSLQP---SVLMQL---------K 181
           |::|  |                | : :  :    |  |   |  ||  ||         :
Db     266 PYMAPERIDPSASRQGYDVRSDVWSLGITLYELATGRFPYPKWNSVFDQLTQVVKGDPPQ 325

Qy     182 LSDGSAYRF--------------EVPTAKFQELRYSVALVLKEMADLEKRC 218
           ||:     |              |  |::||    ::: |   :|  |
Db     326 LSNSEEREFSPSFINFVNLCLTKDESKRPKYKELLKHPFILMYEERTVEVAC 377


RESULT   5
UL70_MCMVS
ID   UL70_MCMVS       STANDARD;       PRT;   964 AA.
AC   Q69153;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   HELICASE/PRIMASE COMPLEX PROTEIN (PROBABLE DNA REPLICATION PROTEIN
DE   UL70).
GN   UL70.
OS   Murine cytomegalovirus (strain Smith).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Betaherpesvirinae; Muromegalovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96082764.
RA   Messerle M., Rapp M., Lucin P., Koszinowski U.H.;
RT   "Characterization of a conserved gene block in the murine
RT   cytomegalovirus genome.";
RL   Virus Genes 10:73-80(1995).
CC   -!- FUNCTION: INVOLVED IN DNA REPLICATION (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL52,
CC       EHV-1 7, EBV BSLF1, HVS-1 56, VZV 6, HCMV AND MCMV UL70.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L07319; AAA96665.1; -.
KW   Helicase; DNA replication.
SQ   SEQUENCE   964 AA;  109704 MW;  70605300E1D85864 CRC64;

  Query Match                    7.3%;  Score 82.5;  DB 1;  Length 964;
  Best Local Similarity   27.4%;  Pred. No. 9.7;
  Matches   51;  Conservative   21;  Mismatches   67;  Indels   47;  Gaps   11;

Qy      30 PPEVAA-----MARLLGDLDRSTF--RKLL-KFVVSSLQGEDCRDGVQRLGVSANLP--- 79
           ||  ||     : || | ||  :    | |: |    |  ||    |:   | |  :|
Db     641 PPRTAAIGAQTLKRLAGILDHTLCLDRDLVCKLNAISHPGECFDTGIYSHGRSIRMPLMY 700

Qy      79 --EEQLGALLAGMHTLLQQALRLPPTSLKPDTFRDQ-----LQELCIPQDLVGDLASVVF 131
             :|  | :|   |:          || |  : |  :||:      ||:|| ||:|
Db     701 KLDEASGLIL---HS------RLNPIFIVPAGYRDRPAEFVLQQLC-PQNL-------TH 743

Qy     132 GSQRPLLDSVAQQQGAWLPHVAD----------FRWRVDVAISTSALARSLQPSVLMQLK 181
           : |  |  | |    : |: |                : |   |:|:  |   | | :   |
Db     744 HGRPPQRDGSADQLTEVVLHITDRACADSDGNFLQSRARRAMSSGRL--PLGPLLRAHLS 801

Qy     182 LSDGSA 187
           | | :
Db     802 LESGQS 807


RESULT   6
MPK4_HUMAN
ID   MPK4_HUMAN      STANDARD;       PRT;   399 AA.
AC   P45985;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   DUAL SPECIFICITY MITOGEN-ACTIVATED PROTEIN KINASE KINASE 4
DE   (EC 2.7.1.-) (MAP KINASE KINASE 4) (JNK ACTIVATING KINASE 1) (C-JUN N-
DE   TERMINAL KINASE KINASE 4) (JNK ACTIVATING KINASE 1) (MAP2K4 OR PRKMK4
GN   OR JNKK1 OR MKK4 OR SERK1).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95232504.
RA   Lin A., Minden A., Martinetto H., Claret F.-X., Lange-Carter C.,
RA   Mercurio F., Johnson G.L., Karin M.;
RT   "Identification of a dual specificity kinase that activates the Jun
RT   kinases and p38-Mpk2.";
RL   Science 268:286-290(1995).
RN   [2]
RP   SEQUENCE FROM N.A
RC   TISSUE=BRAIN,
```

```
RX   MEDLINE; 95141073.
RA   Derijard B., Raingeaud J., Barrett T., Wu I.-H., Han J.,
RA   Ulevitch R.J., Davis R.J.;
RT   "Independent human MAP-kinase signal transduction pathways defined by
RT   MEK and MKK isoforms.";
RL   Science 267:682-685(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98283389.
RA   Su G.H., Hilgers W., Shekher M.C., Tang D.J., Yeo C.J., Hruban R.H.,
RA   Kern S.E.;
RT   "Alterations in pancreatic, biliary, and breast carcinomas support
RT   MKK4 as a genetically targeted tumor suppressor gene.";
RL   Cancer Res. 58:2339-2342(1998).
CC   -!- FUNCTION: DUAL SPECIFICITY KINASE THAT ACTIVATES THE JUN KINASES
CC       MAPK8 (JNK1) AND MAPK9 (JNK2) AS WELL AS MAPK14 (P38) BUT NOT
CC       MAPK1 (ERK2) OR MAPK3 (ERK1).
CC   -!- PTM: ACTIVATED BY PHOSHORYLATION ON SER/THR BY MAP KINASE KINASE
CC       KINASES (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: ABUNDANT EXPRESSION IS SEEN IN THE SKELETAL
CC       MUSCLE. IT IS ALSO WIDELY EXPRESSED IN OTHER TISSUES.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC       MAP KINASE KINASE SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U17743; AAC50127.1; -.
DR   EMBL; L36870; AAC41719.1; -.
DR   EMBL; AF070090; AAC24130.1; -.
DR   EMBL; AF070080; AAC24130.1; JOINED.
DR   EMBL; AF070081; AAC24130.1; JOINED.
DR   EMBL; AF070082; AAC24130.1; JOINED.
DR   EMBL; AF070083; AAC24130.1; JOINED.
DR   EMBL; AF070084; AAC24130.1; JOINED.
DR   EMBL; AF070085; AAC24130.1; JOINED.
DR   EMBL; AF070086; AAC24130.1; JOINED.
DR   EMBL; AF070087; AAC24130.1; JOINED.
DR   EMBL; AF070088; AAC24130.1; JOINED.
DR   EMBL; AF070089; AAC24130.1; JOINED.
DR   HSSP; P11362; 1FGK.
DR   MIM; 601335; -.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PFAM; PF00069; pkinase; 1.
KW   Transferase; Serine/threonine-protein kinase; Tyrosine-protein kinase;
KW   ATP-binding; Phosphorylation.
FT   DOMAIN      5    19      GLY/SER-RICH.
FT   DOMAIN      102  367     PROTEIN KINASE.
FT   NP_BIND     105  113     ATP (BY SIMILARITY).
FT   BINDING     131  131     ATP (BY SIMILARITY).
FT   ACT_SITE    229  229     BY SIMILARITY.
SQ   SEQUENCE   399 AA;  44287 MW;  A472537F2F26770B CRC64;
```

Query Match 7.2%; Score 82; DB 1; Length 399;
Best Local Similarity 20.2%; Pred. No. 3.6;
Matches 59; Conservative 44; Mismatches 89; Indels 100; Gaps 13;

```
Qy       3 AVGAATPYLHHPGDSHSGRVSFLGAQLPPEVAAMARLLGDLDRSTFRKLLKFVVSSLQGE 62
           | |:    :| |    ||           :: |: |:   :|    ::||   :   ::
Db     112 AYGSVNKMVHKP----SG----------QIMAVKRIRSTVDEKEQKQLLMDLDVVMRSS 156

Qy      63 DC----------RDG-----VQRLGVSAN-----------LPEEQLGAL-LAGMHTL 92
           ||          |:|     :: :  | :           :||| || : || :  |
Db     157 DCPYIVQFYGALFREGDCWICMELMSTSFDKFYKYVYSVLDDVIPEEILGKITLATVKAL 216
```

```
Qy      93 --LQQALRLPPTSLKPDTFRDQLQELCIPQDLVGDLASVVFGSQRPLLDSVAQQQGAWL- 150
             |:: |::   :||          |  |  |::   ||    |:||:|: : |
Db     217 NHLKENLKIIHRDIKPSN---------ILLDRSGNIKLCDFGISGQLVDSIAKTRDAGCR 267

Qy     150 PHVADFR---------------WRVDVAISTSALARSLQP---SVLMQL---------K 181
           |::|  |               | : : :  |  |  |  ||  ||          :
Db     268 PYMAPERIDPSASRQGYDVRSDVWSLGITLYELATGRFPYPKWNSVFDQLTQVVKGDPPQ 327

Qy     182 LSDGSAYRF--------------EVPTAKFQELRYSVALVLKEMADLEKRC 218
           ||:   |                |  |::||    ::: |   :|  |
Db     328 LSNSEERFSPSFINFVNLCLTKDESKRPKYKELLKHPFILMYEERAVEVAC 379
```

```
RESULT   7
BIOF_ECOLI
ID   BIOF_ECOLI     STANDARD;      PRT;   384 AA.
AC   P12998;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   8-AMINO-7-OXONONANOATE SYNTHASE (EC 2.3.1.47) (7-KETO-8-AMINO-
DE   PELARGONIC ACID SYNTHETASE) (7-KAP SYNTHETASE) (L-ALANINE--PIMELYL COA
DE   LIGASE).
GN   BIOF.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 89066784.
RA   Otsuka A.J., Buoncristiani M.R., Howard P.K., Flamm J., Johnson O.,
RA   Yamamoto R., Uchida K., Cook C., Ruppert J., Matsuzaki J.;
RT   "The Escherichia coli biotin biosynthetic enzyme sequences predicted
RT   from the nucleotide sequence of the bio operon.";
RL   J. Biol. Chem. 263:19577-19585(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Pearson B.M., McKee R.A.;
RT   "Genetic material for expression of biotin synthetase enzymes.";
RL   Patent number GB2216530, 11-OCT-1989.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE; 97426617.
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner F.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
CC   -!- CATALYTIC ACTIVITY: 6-CARBOXYHEXANOYL-COA + L-ALANINE =
CC       8-AMINO-7-OXONONANOATE + COA + CO(2).
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC   -!- PATHWAY: BIOTIN BIOSYNTHESIS PATHWAY.
CC   -!- SIMILARITY: BELONGS TO CLASS-II OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC       AMINOTRANSFERASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J04423; AAA23516.1; -.
DR   EMBL; A11542; CAA00968.1; -.
DR   EMBL; AE000180; AAC73863.1; -.
DR   PIR; D32025; SYECKP.
DR   ECOGENE; EG10121; BIOF.
DR   PROSITE; PS00599; AA_TRANSFER_CLASS_2; 1.
DR   PFAM; PF00222; aminotran_2; 1.
KW   Biotin biosynthesis; Transferase; Pyridoxal phosphate.
```

```
FT   BINDING     236    236       PYRIDOXAL PHOSPHATE (PROBABLE).
FT   CONFLICT    188    189       AE -> R (IN REF. 2).
SQ   SEQUENCE   384 AA;  41594 MW;  D1AF5C054A5B4B06 CRC64;


  Query Match             7.2%;  Score 81.5;  DB 1;  Length 384;
  Best Local Similarity   19.9%;  Pred. No. 3.8;
  Matches   56;  Conservative   32;  Mismatches   90;  Indels  103;  Gaps   12;

Qy       6 AATPYL---HHP--------------GDSHSGRVSFLGAQLPPEVAAMARLLGDLDRST 47
           ::  ||   |||              |   || ||          :|  ||
Db      45 SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLG------ 99

Qy      48 FRKLLKF---------VVSSLQGEDCRDGVQRLGVSANLPEEQLGALLAGMHTLLQQALR 98
           : : | |         |::::  :: |    ||                  |  | :|
Db      99 YSRALLFISGFAANQAVIAAMMAKEDRIAADRLS----------------HASLLEAAS 141

Qy      99 LPPTSLKPDTFRD------QLQELCIPQDLVGDLASVVF---GSQRPL--LDSVAQQQGA 147
           | |: |:     |       |   | | :|  :   ||   |   ||  :  | ||
Db     142 LSPSQLRRFAHNDVTHLARLLASPCPGQQMV--VTEGVFSMDGDSAPLAEIQQVTQQHNG 199

Qy     148 WLPHVADFRWRVDVAISTSALA---------RSLQPSVLM-------------------- 179
           ||          || | |  :           : ::| :|:
Db     200 WL--------MVDDAHGTGVIGEQGRGSCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTV 251

Qy     179 ---QLKLSDGSAYRFEVPTAKFQELRYSVALVLKEMADLEK 216
             |: :    |   :| |: | || |:|::  :  |  :
Db     252 ADYLLQFARHLIYSTSMPPAQAQALRASLAVIRSDEGDARR 292


RESULT   8
P152_YEAST
ID   P152_YEAST     STANDARD;      PRT;  1337 AA.
AC   P39685;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   NUCLEAR ENVELOPE PORE MEMBRANE PROTEIN POM152 (P150).
GN   POM152 OR YMR129W OR YM9553.05.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=W303;
RX   MEDLINE; 94186543.
RA   Wozniak R.W., Blobel G., Rout M.P.;
RT   "POM152 is an integral protein of the pore membrane domain of the
RT   yeast nuclear envelope.";
RL   J. Cell Biol. 125:31-42(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Badcock K., Churcher C., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL   Submitted (MAR-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: COMPONENT OF THE NUCLEAR PORE COMPLEX WHICH MAY PROVIDE
CC       THE NECESSARY ASYMMETRY REQUIRED FOR ANCHORING STRUCTURES SUCH AS
CC       CYTOPLASMIC FILAMENTS AND THE NUCLEOPLASMIC CAGE.
CC   -!- SUBUNIT: INTERACTS WITH NUP170 AND/OR NUP188.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. NUCLEAR PORE
CC       MEMBRANE.
CC   -!- PTM: THE N-TERMINUS IS BLOCKED.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z31592; CAA83469.1; -.
DR   EMBL; Z48622; CAA88554.1; -.
DR   PIR; A53824; A53824.
DR   SGD; L0001464; POM152.
KW   Nuclear protein; Transmembrane; Repeat; Glycoprotein.
FT   DOMAIN        1    175       PORE SIDE (POTENTIAL).
FT   TRANSMEM    176    195       POTENTIAL.
FT   DOMAIN      196   1337       CISTERNAL SIDE (POTENTIAL).
FT   CARBOHYD    280    280
FT   DOMAIN      390   1276       8 X 24 AA APPROXIMATE REPEATS.
FT   REPEAT      390    413       1.
FT   REPEAT      626    650       2.
FT   REPEAT      732    755       3.
FT   REPEAT      836    859       4.
FT   REPEAT      943    966       5.
FT   REPEAT     1058   1077       6.
FT   REPEAT     1157   1178       7.
FT   REPEAT     1253   1276       8.
SQ   SEQUENCE   1337 AA;  151651 MW;  A024A42069193898 CRC64;


  Query Match             7.2%;  Score 81.5;  DB 1;  Length 1337;
  Best Local Similarity   25.9%;  Pred. No. 18;
  Matches   45;  Conservative   28;  Mismatches   72;  Indels   29;  Gaps    9;

Qy      56 VSSLQGEDCRDGVQRLGVSANLPEEQLGA---LLAGMHTLLQQALRLPPTSLKPDTFRDQ 112
           ::|: |    |  : :| | |:  |  |      ||    :  |      :||:|: |
Db     380 IASITGTGSND--RCIGDSDNVSFEIQGVPPMKLAYSKIVNGQTFSYVDSSLQPEYFESP 437

Qy     113 LQELCIPQDLV-GDLASVVFGSQRPL---LDSVAQQQGAWLPHVADFRWRVDVAISTSAL 168
           ||     |  |:|  : :|  :|:   |||   | |        | :::|    |  |
Db     438 LQSSKSKQSFTQGELNDLKWGRNQPVNINLDSSITQDG-------KFAYKIDKI--TDGL 488

Qy     169 ARSLQ-PSVLMQLKLSDGSAYRF---EVPTAKFQE-------LRYSVALVLKEM 211
           :   |:  :||     :| |   ||| |  :|        : |:|:| :|:
Db     489 GNVVDFTSLPEELKKRYDLSYNFNVHEVPRAALEERFDPKSPTKRSIAIVFEEI 542


RESULT   9
CH60_NEIME
ID   CH60_NEIME     STANDARD;      PRT;   544 AA.
AC   P42385;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   60 KD CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN) (63 KD STRESS
DE   PROTEIN) (GSP63).
GN   MOPA OR GROEL OR HSP63.
OS   Neisseria meningitidis.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=2996;
RX   MEDLINE; 95264914.
RA   Pannekoek Y., Dankert J., van Putten J.P.M.;
RT   "Construction of recombinant neisserial Hsp60 proteins and mapping of
RT   antigenic domains.";
RL   Mol. Microbiol. 15:277-285(1995).
CC   -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC       PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC       CONDITIONS (BY SIMILARITY).
CC   -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC       7 SUBUNITS (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND LOOSELY ASSOCIATED WITH THE
CC       OUTER MEMBRANE.
CC   -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; Z22956; CAA80532.1; -.
DR   HSSP; P06139; 1GRL.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
DR   PFAM; PF00118; cpn60_TCP1; 1.
KW   Chaperone; ATP-binding; Antigen.
SQ   SEQUENCE   544 AA;  57580 MW;  B0789DA091149D57 CRC64;


  Query Match             7.2%;  Score 81;  DB 1;  Length 544;
  Best Local Similarity   19.9%;  Pred. No. 6.5;
  Matches   61;  Conservative   45;  Mismatches   84;  Indels  116;  Gaps   14;

Qy      25 LGAQLPPEVAAMAR-LLGD-------LDRSTFRKLLKFVVSSLQGEDCRDGVQRL----- 72
           :|||:  |||:    : ||       | :|   : :|:| : :   | : |: :
Db      69 MGAQMVKEVASKTNDVAGDGTTTATVLAQSIVAEGMKYVTAGMNPTDLKRGIDKAVAALV 128

Qy      72 -------------------GVSANLPEEQLGALLA---------GMHTL---------- 93
                              :|||  :||:||::|          |: |:
Db     129 EELKNIAKPCDTSKEIAQVGSISAN-SDEQVGAIIAEAMEKVGKEGVITVEDGKSLENEL 187

Qy      93 -LQQALRLPPTSLKP------------------------DTFRDQLQE----------- 116
            : : ::     | |                        || |
Db     188 DVVEGMQFDRGYLSPYFINDAEKQIAGLDNPFVLLFEKKISNIRDLLPVLEKVAKASRPL 247

Qy     116 LCIPQDLVGD-LASVVFGSQRPLLDSVAQQQGAWLPHVADFRWRV--DVAISTSALARSL 172
           | | ::: |: ||::|  : | :| :|| :     |   | |  :  |:|| |     |
Db     248 LIIAENVEGEALATLVVNNIRGILKTVAVK----APGFGDRRKAMLQDIAILTGGTVISE 303

Qy     173 QPSVLMQLKLSD--GSAYRFEVP---------------TAKFQELRYSVALVLKEMADL 214
           :  : ::     |  | | | |:                |:| |:|  :     :  |
Db     304 EVGLFLEKATLDNLGQAKRIEIGKENTTIIDGFGDAAQIEARFAEIRQQIETATSDY-DK 362

Qy     215 EKRCER 220
           ||  ||
Db     363 EKLQER 368


RESULT  10
PFES_PSEAE
ID   PFES_PSEAE     STANDARD;      PRT;   446 AA.
AC   Q04804;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   SENSOR PROTEIN PFES (EC 2.7.3.-).
GN   PFES.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC   Pseudomonas.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PAO;
RX   MEDLINE; 93368425.
RA   Dean C.R., Poole K.;
RT   "Expression of the ferric enterobactin receptor (PfeA) of Pseudomonas
RT   aeruginosa: involvement of a two-component regulatory system.";
RL   Mol. Microbiol. 8:1095-1103(1993).
CC   -!- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM PFER/PFES.
CC       MAY ACTIVATE PFER BY PHOSPHORYLATION.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC       (POTENTIAL).
CC   -!- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE
CC       KINASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L07739; AAA25930.1; -.
DR   PFAM; PF00512; signal; 1.
DR   PFAM; PF00672; DUF5; 1.
KW   Sensory transduction; Transferase; Kinase; Phosphorylation;
KW   Transmembrane; Inner membrane.
FT   DOMAIN        1      9       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     10     30       POTENTIAL.
FT   DOMAIN       31    155       PERIPLASMIC (POTENTIAL).
FT   TRANSMEM    156    176       POTENTIAL.
FT   DOMAIN      177    446       CYTOPLASMIC (POTENTIAL).
FT   MOD_RES     244    244       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ   SEQUENCE   446 AA;  50597 MW;  62B78FED1B4FE73E CRC64;


  Query Match             7.1%;  Score 80.5;  DB 1;  Length 446;
  Best Local Similarity   24.3%;  Pred. No. 5.6;
  Matches   46;  Conservative   18;  Mismatches   62;  Indels   63;  Gaps    8;

Qy      27 AQLPP---------EVAAMARLLGD-LDRS----------TFRKLLKFVVSSLQGEDCRD 66
           ::|||          |:  | ||| | || :          |   |   |  :|: : | :
Db     259 SELPPEQLRQRLDREIGDMQRLLEDTLDLAWMDTERPQLPTEPVLELSVWEALRDDACFE 318

Qy      67 G--------VQRLGVSANLPEEQLGALLAGMHTLLQQALRLPP----TSLK---------- 106
                   ||||   : |  | :|   |  ||: |:|  |     ||
Db     319 SGWDPARLPCRLGVDCRV-EVHLDSLAQAMENLLRNAIRHSPEDGTVSLDGEREGDFWHL 377

Qy     106 ------PDTFRDQLQELCIPQDLVGDLASVVFGSQRPLLDSVAQQQGA------------ 148
                 |    |||: : :|   : | |   ||    :     : ||
Db     378 RLQDQGPGVAEDQLERIFLPYQRLDDSAGEGFGLGLAIARRAIELQGGRLWASNGKPGLC 437

Qy     148 ---WLPHVA 153
              |||  |
Db     438 LHLWLPAAA 446


RESULT  11
BGLG_ECOLI
ID   BGLG_ECOLI     STANDARD;      PRT;   278 AA.
AC   P11989;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   CRYPTIC BETA-GLUCOSIDE BGL OPERON ANTITERMINATOR.
GN   BGLG OR BGLC.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE; 87222180.
RA   Schnetz K., Toloczyki C., Rak B.;
RT   "Beta-glucoside (bgl) operon of Escherichia coli K-12: nucleotide
RT   sequence, genetic organization, and possible evolutionary
RT   relationship to regulatory components of two Bacillus subtilis
RT   genes.";
RL   J. Bacteriol. 169:2579-2590(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE; 93315143.
RA   Burland V.D., Plunkett G. III, Daniels D.L., Blattner F.R.;
RT   "DNA sequence and analysis of 136 kilobases of the Escherichia coli
RT   genome: organizational symmetry around the origin of replication.";
RL   Genomics 16:551-561(1993).
RN   [3]
RP   SEQUENCE OF 1-197 FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE; 87273510.
RA   Mahadevan S., Wright A.;
RT   "A bacterial gene involved in transcription antitermination:
RT   regulation at a rho-independent terminator in the bgl operon of E.
```

```
RT   coli.";
RL   Cell 50:485-494(1987).
RN   [4]
RP   REGULATION BY PHOSPHORYLATION.
RX   MEDLINE; 90341774.
RA   Amster-Choder O., Wright A.;
RT   "Regulation of activity of a transcriptional anti-terminator in E.
RT   coli by phosphorylation in vivo.";
RL   Science 249:540-542(1990).
RN   [5]
RP   REGULATION BY PHOSPHORYLATION.
RX   MEDLINE; 90311345.
RA   Schnetz K., Rak B.;
RT   "Beta-glucoside permease represses the bgl operon of Escherichia coli
RT   by phosphorylation of the antiterminator protein and also interacts
RT   with glucose-specific enzyme III, the key element in catabolite
RT   control.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:5074-5078(1990).
RN   [6]
RP   RNA-BINDING.
RX   MEDLINE; 90381772.
RA   Houman F., Diaz-Torre M.R., Wright A.;
RT   "Transcriptional antitermination in the bgl operon of E. coli is
RT   modulated by a specific RNA binding protein.";
RL   Cell 62:1153-1163(1990).
CC   -!- FUNCTION: MEDIATES THE POSITIVE REGULATION OF THE BETA-GLUCOSIDE
CC       (BGL) OPERON BY FUNCTIONING AS A TRANSCRIPTIONAL ANTITERMINATOR.
CC       THIS IS A RNA-BINDING PROTEIN THAT RECOGNIZES A SPECIFIC SEQUENCE
CC       LOCATED JUST UPSTREAM OF TWO TERMINATION SITES WITHIN THE OPERON.
CC   -!- PTM: PHOSPHORYLATED AND INACTIVATED BY BGLF (II-BGL). THE DEGREE
CC       OF PHOSPHORYLATION IS DEPENDENT ON THE PRESENCE OR ABSENCE OF
CC       BETA-GLUCOSIDES WHICH ACT AS INDUCERS OF THE OPERON EXPRESSION.
CC       ADDITION OF INDUCER RESULT IN THE RAPID DEPHOSPHORYLATION OF BGLG.
CC   -!- SIMILARITY: BELONGS TO THE TRANSCRIPTIONAL ANTITERMINATOR BGLG
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M16487; AAA23509.1; -.
DR   EMBL; L10328; AAA62074.1; -.
DR   EMBL; AE000449; AAC76746.1; -.
DR   EMBL; M17098; AAA23512.1; -.
DR   PIR; A27161; A27161.
DR   PIR; B25977; B25977.
DR   HSSP; P15401; 1AUU.
DR   ECOGENE; EG10116; BGLG.
DR   PROSITE; PS00654; ANTITERMINATORS_BGLG; 1.
DR   PFAM; PF00874; BglG_antitermin; 2.
KW   Transcription regulation; Activator; RNA-binding; Phosphorylation.
SQ   SEQUENCE   278 AA;  32097 MW;  5ACF1A14BF438B4F CRC64;


  Query Match             7.1%;  Score 80;  DB 1;  Length 278;
  Best Local Similarity   21.8%;  Pred. No. 3.5;
  Matches   51;  Conservative   42;  Mismatches   91;  Indels   50;  Gaps    8;

Qy      17 SH--SGRVSFLGAQLPPEVAAMARLLGDLDRSTFRKLLKFVVSSLQGEDCRDGVQRLGVS 74
           ||  :||:| | : :| || |    :  | :    ||   :  ||  : |:  ::|   :
Db      54 SHELNGRLSELLSHIPLEVMATCDRIISLAQERLGKLQDSIYISLT-DHCQFAIKRFQQN 112

Qy      75 ANLPEEQLGALLAGMHTLLQQALRLPPTSLKPDTFRDQLQELCIPQDLVGDLASVVFGSQ 134
             ||      ||  :  |  :  :|   :|   |  |:   : :|:| || :|  :  :|
Db     113 VLLP----NPLLWDIQRLYPKEFQLGEEAL---TIIDKRLGVQLPKDEVGFIAMHLVSAQ 165

Qy     135 --------------RPLL------------DSVAQQQGAWLPHVADFRWRVDVAISTS 166
                        | :|  .           :|:: |:   : |:    ||:    | :
Db     166 MSGNMEDVAGVTQLMREMLQLIKFQFSLNYQEESLSYQR--LVTHLKFLSWRILEHASIN 223

Qy     167 ALARSLQPSVLM----------QLKLSDGSAYRFEVPTAKFQELRYSVALVLKE 210
                 ||| :|            :: :  |  |: ::  |:  |  ::  | ||
Db     224 DSDESLQQAVKQNYPQAWQCAERIAIFIGLQYQRKISPAEIMFLAINIERVRKE 277


RESULT  12
MYSC_CAEEL
ID   MYSC_CAEEL     STANDARD;      PRT;  1947 AA.
AC   P12845;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   MYOSIN HEAVY CHAIN C (MHC C).
GN   MYO-2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 89178677.
RA   Dibb N.J., Maruyama I.N., Krause M., Karn J.;
RT   "Sequence analysis of the complete Caenorhabditis elegans myosin
RT   heavy chain gene family.";
RL   J. Mol. Biol. 205:603-613(1989).
RN   [2]
RP   SEQUENCE OF 1-21; 112-371 AND 1501-1772 FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 85201409.
RA   Karn J., Dibb N.J., Miller D.M.;
RT   "Cloning nematode myosin genes.";
RL   Cell Muscle Motil. 6:185-237(1985).
CC   -!- FUNCTION: MUSCLE CONTRACTION.
CC   -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC       HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC       AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC   -!- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.
CC   -!- TISSUE SPECIFICITY: FOUND EXCLUSIVELY IN THE PHARYNGEAL MUSCLE.
CC   -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC       CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC       CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC   -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC       ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC   -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC       MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
CC       SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC       SUBFRAGMENT (S2).
CC   -!- MISCELLANEOUS: THERE ARE FOUR DIFFERENT MYOSIN HEAVY CHAINS IN C.
CC       ELEGANS.
CC   -!- SIMILARITY: THE GLOBULAR HEAD AA SEQUENCE SHOWS A HIGH DEGREE OF
CC       SIMILARITY WITH THE GLOBULAR HEAD SEQUENCES OF MUSCLE & NONMUSCLE
CC       HEAVY CHAINS. BY CONTRAST THE ROD SEQUENCE IS LESS CONSERVED, BUT
CC       THE PERIODICITIES OF HYDROPHOBIC & CHARGED RESIDUES, WHICH DICTATE
CC       THE ALPHA-HELICAL COILED-COIL STRUCTURE, ARE CONSERVED.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X08066; CAA30855.1; -.
DR   EMBL; M37233; AAA28121.1; -.
DR   EMBL; M37235; AAA28122.1; -.
DR   EMBL; M37236; AAA28123.1; -.
DR   PIR; S05697; S05697.
DR   HSSP; P08799; 1MND.
DR   PFAM; PF00063; myosin_head; 1.
DR   PFAM; PF01576; Myosin_tail; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
```

```
KW   Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW   ATP-binding; Methylation; Alkylation; Heptad repeat pattern;
KW   Multigene family.
FT   DOMAIN        1    855       GLOBULAR HEAD (S1).
FT   DOMAIN      856   1947       RODLIKE TAIL (S2 AND LMM DOMAINS).
FT   DOMAIN      856   1947       COILED COIL (POTENTIAL).
FT   NP_BIND     174    181       ATP (BY SIMILARITY).
FT   DOMAIN      668    690       ACTIN-BINDING.
FT   DOMAIN      773    787       ACTIN-BINDING.
FT   MOD_RES     125    125       METHYLATION (TRI-) (POTENTIAL).
FT   MOD_RES     708    708       ALKYLATION (SH-1).
FT   MOD_RES     718    718       ALKYLATION (SH-2).
FT   CONFLICT    132    132       E -> D (IN REF. 2).
FT   CONFLICT    137    137       M -> I (IN REF. 2).
SQ   SEQUENCE   1947 AA;  223009 MW;  6D65AE99BD013627 CRC64;


  Query Match             7.1%;  Score 80;  DB 1;  Length 1947;
  Best Local Similarity   22.2%;  Pred. No. 39;
  Matches   57;  Conservative   41;  Mismatches   97;  Indels   62;  Gaps   10;

Qy      21 RVSFLGAQLPPE-------VAAMARLLG--DLDRSTFRKLLKFV-VSSLQGEDCRDGVQR 70
           |  |    | |:       :||   ::|  |  : :   | : |    |: |: | |: :
Db     719 RKGFPNRTLHPDFVQRYALLAADESIIGKTDAKKGSALMLARLVKEKKLEEENFRVGLTK 778

Qy      71 L----GVSANLPEEQLGALLAGMHTLLQQALRLPPTSLKPDTFRDQLQELCIPQDLVGDL 126
           :    |: |:| |:     || : | ||  :|    :::     :::  | | |  :
Db     779 VFFKAGIVAHL-EDLRDQSLAQLITGLQAQIRWYYQTIERKRRVEKITALKIIQRNIRSW 837

Qy     127 ASV-------VFGSQRPLLDS--VAQQQGAWLPHVADFRWRVDVAISTSALARSLQ---- 174
           | :       ::|  :||::|  :  |       ||  :   |  :          | ||
Db     838 AELRTWVWFKLYGKVKPLVNSGKIEAQYEKLQETVATLK---DTVVQEEEKKRQLQEGAE 894

Qy     174 ------PSVLMQLKLSDGS-------------------------AYRFEVPTAKFQELRY 202
                 :| ||: | ||                          : : ||  |:  |:
Db     895 RLNKETADLLAQLEASKGSTREVEERMTAMNEQKVALEGKLADASKKLEVEEARAVEINK 954

Qy     203 SVALVLKEMADLEKRCE 219
           ||  | |||:| |:
Db     955 QKKLVEAECADLKKNCQ 971


RESULT  13
GLND_ECOLI
ID   GLND_ECOLI     STANDARD;      PRT;   890 AA.
AC   P27249;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   [PROTEIN-PII] URIDYLYLTRANSFERASE (EC 2.7.7.59) (PII URIDYLYL-
DE   TRANSFERASE) (URIDYLYL REMOVING ENZYME) (UTASE).
GN   GLND.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RA   Park S.-C., Kim I.H., Rhee S.G.;
RL   Submitted (XXX-1992) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / W3110;
RX   MEDLINE; 94018640.
RA   van Heeswijk W.C., Rabenberg M., Westerhoff H.V., Kahn D.;
RT   "The genes of the glutamine synthetase adenylylation cascade are not
RT   regulated by nitrogen in Escherichia coli.";
RL   Mol. Microbiol. 9:443-458(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / W3110;
RX   MEDLINE; 94261430.
RA   Fujita N., Mori H., Yura T., Ishihama A.;
RT   "Systematic sequencing of the Escherichia coli genome: analysis of
RT   the 2.4-4.1 min (110,917-193,643 bp) region.";
RL   Nucleic Acids Res. 22:1637-1639(1994).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE; 97426617.
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner F.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [5]
RP   SEQUENCE FROM N.A.
RA   Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,
RA   Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA   Lashkari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,
RA   Davis R.W.;
RL   Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: MODIFIES, BY URIDYLYLATION OR DEURIDYLYLATION THE PII
CC       (GLNB) REGULATORY PROTEIN.
CC   -!- CATALYTIC ACTIVITY: UTP + [PROTEIN-PII] = PYROPHOSPHATE +
CC       URIDYLYL-[PROTEIN-PII].
CC   -!- SIMILARITY: TO OTHER BACTERIAL GLND.
CC   -!- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 703
CC       ONWARD AND IS SHORTER (727 AA) DUE TO A FRAMESHIFT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M96431; AAA23878.1; -.
DR   EMBL; Z21842; CAA79887.1; -.
DR   EMBL; D26562; CAB20283.1; ALT_FRAME.
DR   EMBL; AE000126; AAC73278.1; -.
DR   EMBL; U70214; AAB08596.1; -.
DR   PIR; S31962; S31962.
DR   PIR; S36254; S36254.
DR   ECOGENE; EG11411; GLND.
KW   Transferase; Nucleotidyltransferase.
FT   CONFLICT     47     47       L -> S (IN REF. 1).
FT   CONFLICT    225    225       G -> A (IN REF. 3).
FT   CONFLICT    357    358       RG -> TR (IN REF. 1).
FT   CONFLICT    523    523       D -> H (IN REF. 1).
FT   CONFLICT    703    703       A -> R (IN REF. 2).
SQ   SEQUENCE   890 AA;  102390 MW;  6FF0006341A71D34 CRC64;


  Query Match             7.0%;  Score 79.5;  DB 1;  Length 890;
  Best Local Similarity   25.8%;  Pred. No. 16;
  Matches   56;  Conservative   25;  Mismatches   77;  Indels   59;  Gaps   11;

Qy      11 LHHPGDSHSG-----RVSFLGAQLPPEVAAMARLLGDL------DRSTFRKLLKFVVSSL 59
           | |||  . |     | | ||| :|     : ::  ||      |  | | :||  : |
Db     425 LRHPGAVRRGLLPMHRHSVLGAYMPQWSHIVGQMQFDLFHAYTVDEHTIRVMLK--LESF 482

Qy      60 QGEDCRDGVQR----LGVSANLPEEQLGALLAGMHTLLQQALRLPPTSLKPDTFRDQLQE 115
             |: |   ||    : |   ||  :|  : |  | : :                 |
Db     483 ASEETR---QRHPLCVDVWPRLPSTELIFIAALFHDIAKG-------------RGGDHS 525

Qy     116 LCIPQDLVGDLASVVFGSQRPLLDSVAQQQGAWLPHVADFRWRVDVAISTSALARSLQ-P 174
           :   ||       ||  ::   |:|   |  |||         |  : :| :|  | :| |
Db     526 ILGAQD-------VVHFAELHGLNSRETQLVAWLV-------RQHLLMSVTAQRRDIQDP 571

Qy     175 SVLMQLKLSDGSAYRFEVPTAKFQELRYSVALVLKEM 211
           |: |       :  || |     ||| | | : ::
Db     572 EVIKQ--------FAEEVQTE--NRLRYLVCLTVADI 598
```

```
RESULT  14
QA1S_NEUCR
ID   QA1S_NEUCR     STANDARD;      PRT;   918 AA.
AC   P11637;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   01-FEB-1991 (Rel. 17, Last annotation update)
DE   QUINATE REPRESSOR.
GN   QA-1S.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Pyrenomycetes; Sordariales;
OC   Sordariaceae; Neurospora.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=74-OR23-1A;
RX   MEDLINE; 89293848.
RA   Geever R.F., Huiet L., Baum J.A., Tyler B.M., Patel V.B.,
RA   Rutledge B.J., Case M.E., Giles N.H.;
RT   "DNA sequence, organization and regulation of the qa gene cluster of
RT   Neurospora crassa.";
RL   J. Mol. Biol. 207:15-34(1989).
CC   -!- FUNCTION: REPRESSOR FOR ENZYMES AND PROTEINS OF QUINATE
CC       METABOLISM.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X14603; CAA32753.1; -.
DR   PIR; S04255; S04255.
DR   PIR; E31277; E31277.
DR   PFAM; PF01487; DHquinase_I; 1.
DR   PFAM; PF01488; Shikimate_DH; 1.
KW   Quinate metabolism; Transcription regulation; Repressor; DNA-binding.
SQ   SEQUENCE   918 AA;  100580 MW;  67EDA399CBF098B2 CRC64;


  Query Match             7.0%;  Score 79.5;  DB 1;  Length 918;
  Best Local Similarity   27.0%;  Pred. No. 17;
  Matches   40;  Conservative   15;  Mismatches   44;  Indels   49;  Gaps    6;

Qy      17 SHSGRVSFLGAQLPPEVAAMARLLGDLDRSTFRKLLKFVVSSLQGEDCRDGVQRLGVSAN 76
           | ||  | ||   |   : ::|:::|:: | |   ::  ||
Db     349 SESGPTSPLGL-APHRASEISRVVGEIRRDTVIPIILHVV-------------------- 388

Qy      77 LPEEQL--GALLAGMHTLLQQALRLPPTSLKPDTFRDQLQELCIPQDLVGDLASV----- 130
             ||  |   ||||   | |  |||| |  |  |        | :   |:| | :|
Db     388 FPERALYEEALLALYMTYLNHALRLAPDYLTVD--------LGLDSGLLGQLTTVQGTTK 439

Qy     130 VFGSQRPLLDSVAQQQGAWLPHVADFRW 157
           | |:::             |  |   ||
Db     440 VIGNKQ------------LAEVNSPRW 454


RESULT  15
CNRB_BOVIN
ID   CNRB_BOVIN     STANDARD;      PRT;   853 AA.
AC   P23439;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   ROD CGMP-SPECIFIC 3',5'-CYCLIC PHOSPHODIESTERASE BETA-SUBUNIT
DE   (EC 3.1.4.17) (GMP-PDE BETA).
GN   PDE6B OR PDEB.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae;
OC   Bovinae; Bos.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90330632.
RA   Lipkin V.M., Khramtsov N.V., Vasilevskaya I.A., Atabekova N.V.,
RA   Muradov K.G., Gubanov V.V., Li T., Johnston J.P., Volpp K.J.,
RA   Applebury M.L.;
RT   "Beta-subunit of bovine rod photoreceptor cGMP phosphodiesterase.
RT   Comparison with the phosphodiesterase family.";
RL   J. Biol. Chem. 265:12955-12959(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90267525.
RA   Lipkin V.M., Gubanov V.V., Khramtsov N.V., Vasilevskaya I.A.,
RA   Atabekova N.V., Muradov K.G., Shuvaeva T.M., Surina E.A.,
RA   Zagranichny V.E., Li T.;
RT   "Cyclic GMP phosphodiesterase from bovine retina. Amino acid sequence
RT   of beta-subunit and nucleotide sequence of corresponding cDNA.";
RL   Bioorg. Khim. 16:118-120(1990).
CC   -!- FUNCTION: THIS PROTEIN PARTICIPATES IN PROCESSES OF
CC       TRANSMISSION AND AMPLIFICATION OF THE VISUAL SIGNAL. NECESSARY
CC       FOR THE FORMATION OF A FUNCTIONAL PHOSPHODIESTERASE HOLOENZYME.
CC   -!- CATALYTIC ACTIVITY: GUANOSINE 3',5'-CYCLIC PHOSPHATE + H(2)O =
CC       GUANOSINE 5'-PHOSPHATE.
CC   -!- SUBUNIT: OLIGOMER COMPOSED OF TWO CATALYTIC CHAINS (ALPHA AND
CC       BETA), AN INHIBITORY CHAIN (GAMMA) AND THE DELTA CHAIN.
CC   -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED.
CC   -!- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE PHOSPHODIESTERASE
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J05553; AAA30440.1; -.
DR   EMBL; X57146; CAA40436.1; -.
DR   PIR; A36617; A36617.
DR   PIR; S19145; S19145.
DR   PRINTS; PR00387; PDIESTERASE1.
DR   PROSITE; PS00126; PDEASE_I; 1.
DR   PFAM; PF00037; fer4; 1.
DR   PFAM; PF00233; PDEase; 1.
DR   PFAM; PF01590; GAF; 2.
KW   Hydrolase; cGMP; Vision; Prenylation; Lipoprotein; Membrane.
FT   CHAIN         1    850       ROD CGMP-SPECIFIC 3',5'-CYCLIC
FT                                PHOSPHODIESTERASE BETA-SUBUNIT.
FT   PROPEP      851    853       REMOVED IN MATURE FORM (BY SIMILARITY).
FT   LIPID       850    850       GERANYL-GERANYL (BY SIMILARITY).
FT   CONFLICT    483    484       DE -> EQ (IN REF. 2).
SQ   SEQUENCE   853 AA;  98330 MW;  C4B3F22CFFE7F2FB CRC64;


  Query Match             7.0%;  Score 79;  DB 1;  Length 853;
  Best Local Similarity   20.7%;  Pred. No. 17;
  Matches   49;  Conservative   36;  Mismatches   66;  Indels   86;  Gaps   10;

Qy      32 EVAAMARLLGD------LDRSTFRKLLKFVVSSLQGEDC-------RDGVQRLGVSANLP 78
           | ||:  |: |       ::|  | |:|: : | |  : |        |:||  |
Db      55 ESAALFELVQDMQENVNMERVVF-KILRRLCSILHADRCSLFMYRQRNGVAEL------- 107

Qy      79 EEQLGALLAGMHTLLQQALRLPPTSLKPDTFRDQLQELCIPQDLVGDLASVVFGSQRPLL 138
                       | ||   |::||:    |::  :| |      : :||     ::
Db     107 ----------------ATRL--FSVQPDSV---LEDCLVPPD-----SEIVFPLDIGVV 139

Qy     139 DSVAQ-------QQGAWLPHVADFRWRVDVAISTSALARSLQ-----PSVLMQLKLSDGS 186
             |||       |     || : |   :   :: : ||  :       :|:| :   ||
Db     140 GHVAQTKKMVNVQDVMECPHFSSFADELTDYVTRNILATPIMNGKDVVAVIMAVNKLDGP 199

Qy     187 -------------------------AYRFEVPTAKFQELRYSVALVLKEMADLEKR 217
                                    :|     | : | | :|   | :|: |:|::
```

Db 200 CFTSEDEDVFLKYLNFGTLNLKIYHLSYLHNCETRRGQVLLWSANKVFEELTDIERQ 256

Search completed: May 15, 2000, 05:34:23
Job time: 12493 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: May 15, 2000, 05:33:25 ; Search time 55.93 Seconds
(without alignments)
277.684 Million cell updates/sec

Title: US-09-223-796-4
Perfect score: 1132
Sequence: 1 MSAVGAATPYLHHPGDSHSG.....ALVLKEMADLEKRCERRLQD 224

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : SPTREMBL_12:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	96	8.5	195	11	Q63829	Q63829 mus musculu
2	93.5	8.3	676	5	Q27512	Q27512 caenorhabdi
3	89.5	7.9	202	4	Q9Y6G5	Q9y6g5 homo sapien
4	88.5	7.8	413	2	Q08859	Q08859 klebsiella
5	86.5	7.6	332	10	O48693	O48693 arabidopsis
6	86.5	7.6	1212	5	Q9XZ29	Q9xz29 drosophila
7	85.5	7.6	2182	11	O88480	O88480 rattus norv
8	83	7.3	218	2	O88031	O88031 streptomyce
9	83	7.3	418	4	Q9Y2F1	Q9y2f1 homo sapien
10	82	7.2	992	11	Q9WU50	Q9wu50 mus musculu
11	82	7.2	1224	2	P95629	P95629 rhizobium m
12	81.5	7.2	800	11	Q62037	Q62037 mus musculu
13	81	7.2	145	1	O28746	O28746 archaeoglob
14	80.5	7.1	161	2	Q9WZJ7	Q9wzj7 thermotoga
15	80	7.1	311	2	Q9ZGG9	Q9zgg9 heliobacill
16	80	7.1	381	2	O85846	O85846 sphingomona
17	80	7.1	1968	5	Q20439	Q20439 caenorhabdi
18	79.5	7.0	831	5	Q94979	Q94979 drosophila
19	79.5	7.0	918	3	Q01393	Q01393 neurospora
20	79.5	7.0	1321	2	O86653	O86653 streptomyce
21	79.5	7.0	1386	4	O75064	O75064 homo sapien
22	78.5	6.9	488	2	Q46979	Q46979 escherichia
23	78.5	6.9	548	2	O51832	O51832 myzus persi
24	78.5	6.9	678	3	O94235	O94235 schizosacch
25	78	6.9	4151	5	O96936	O96936 drosophila
26	77.5	6.8	1218	5	Q23326	Q23326 caenorhabdi
27	77.5	6.8	4834	4	O95714	O95714 homo sapien
28	77	6.8	97	12	Q66002	Q66002 california
29	77	6.8	97	12	Q66003	Q66003 california
30	77	6.8	375	10	P93403	P93403 oryza sativ
31	77	6.8	444	2	Q9X6W7	Q9x6w7 azospirillu
32	77	6.8	666	2	Q9X838	Q9x838 streptomyce
33	77	6.8	1198	2	Q53730	Q53730 streptomyce
34	77	6.8	3603	2	P94459	P94459 bacillus su
35	76.5	6.8	340	2	O32441	O32441 pseudomonas
36	76.5	6.8	436	2	O86484	O86484 streptomyce
37	76.5	6.8	575	5	Q24444	Q24444 drosophila
38	76.5	6.8	776	5	O44708	O44708 drosophila
39	76.5	6.8	2626	11	Q9Z1N3	Q9z1n3 rattus norv
40	76	6.7	433	2	Q9X8V6	Q9x8v6 streptomyce
41	76	6.7	696	12	P89467	P89467 herpes simp
42	76	6.7	947	2	P95022	P95022 mycobacteri
43	76	6.7	1236	2	O06264	O06264 mycobacteri
44	75.5	6.7	292	1	Q9YE72	Q9ye72 aeropyrum p
45	75.5	6.7	361	2	O30854	O30854 vibrio chol

ALIGNMENTS

RESULT 1
Q63829
ID Q63829 PRELIMINARY; PRT; 195 AA.
AC Q63829;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE BUP=5'OF BMI-1 PROVIRAL INSERTION LOCUS.
GN BUP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93165034.
RA HAUPT Y., BARRI G., ADAMS J.M.;
RT "Nucleotide sequence of bup, an upstream gene in the bmi-1 proviral
RT insertion locus.";
RL Mol. Biol. Rep. 17:17-20(1992).
DR EMBL; S54914; AAB25378.1; -.
DR MGD; MGI:88218; Bup.
SQ SEQUENCE 195 AA; 22037 MW; 827A84F1 CRC32;

Query Match 8.5%; Score 96; DB 11; Length 195;
Best Local Similarity 20.8%; Pred. No. 0.2;
Matches 43; Conservative 32; Mismatches 78; Indels 54; Gaps 8;

```
Qy      41 GDLDRSTFRKLLKFVVSSLQGEDCRDGVQRLGVSANLPEEQLGALLAGMHTLLQQALRLP 100
           |  | : |  ||:    ||             : |   |  |       |  |:|   :
Db      18 GSFDSNAFALLLRAAFQSL-------------LDARADEAALD------HPYLKQ---ID 55

Qy     101 PTSLK------------------PDTFRDQLQELCIPQDLVGDLASVVFGSQRPLLDSV 141
           |  ||                  |    |::    :: : :|    : : :  |:::
Db      56 PVVLKHCHAAAATCILEAGHNQVDKSTLSTYLEDCKFDRERI-ELFCTEYQNNKNSLETL 114

Qy     142 AQQQGAWLPHVADFRWRVDVAISTSALARSLQPSVLMQLKL--SDGSAY---RFEVPTAK 196
               |  |||: |  ||::  | |: | :  :|  |: | :  :|  :|    |     :
Db     115 LGSIGRSLPHITDVSWRLEYQIKTNQLHKMYRPGYLVTLNVENNDSQSYPEINFSCNMEQ 174

Qy     197 FQELRYSVALVLKEMADLEKRCERRLQ 223
            |:|       : :: |  |  ||  |
Db     175 LQDL-------VGKLKDASKSLERATQ 194
```

```
RESULT   2
Q27512
ID   Q27512      PRELIMINARY;      PRT;   676 AA.
AC   Q27512;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   NEX-2 PROTEIN.
GN   NEX-2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BERKS M.;
RL   Submitted (JAN-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
CC   -!- FUNCTION: CALCIUM/PHOSPHOLIPID-BINDING PROTEIN WHICH PROMOTES
CC       MEMBRANE FUSION AND IS INVOLVED IN EXOCYTOSIS.
CC   -!- SIMILARITY: TO OTHER PROTEINS OF THE ANNEXIN FAMILY.
DR   EMBL; Z29443; CAA82571.1; -.
DR   HSSP; P79134; 1AVC.
DR   PROSITE; PS00223; ANNEXIN; 1.
DR   PFAM; PF00191; annexin; 4.
DR   PRINTS; PR00196; ANNEXIN.
KW   Annexin; Calcium/phospholipid-binding; Repeat.
SQ   SEQUENCE   676 AA;  74969 MW;  00E36C79 CRC32;


  Query Match             8.3%;  Score 93.5;  DB 5;  Length 676;
  Best Local Similarity   22.7%;  Pred. No. 1.8;
  Matches   41;  Conservative   32;  Mismatches   63;  Indels   45;  Gaps    8;

Qy      43 LDRSTFRKLLKFVVSSLQGEDCRDGVQRLGVSANLPEEQLGALLAGMHTLLQQALRLPPT 102
           |: | |||| : |  |   |               | |||    || ::     : ||
Db      50 LNNSKFRKLCELVKESRSHE-------------MLSEEQ----LATLYDCATTSTRL--- 90

Qy     103 SLKPDTFRDQLQELCIPQDLV--GDLASVVFG---------------SQRPLLDSVAQQQ 145
             |: | |  ::|   ::::  |:|                    ::  |: : :|:
Db      90 ---PNDFADDQEDL---KNIIRYGELFKACHAINSTDFIQKSEDLKDEEKVALERIVEQK 143

Qy     146 GA-WLPHVADFRWRVDVAISTSALARSLQPSVLMQLKLSDGSAYRFEVPTAKFQELRYSV 204
            |    :  |  | |:: :: | :|:||:| : :|:    |    ||     | : |  :
Db     144 LAESAKNEKDIAWNVNIVVANSYVAKSLRPVINIQMPTVGGDT-NFEFDIDSFAQFRQQL 202

Qy     205 A 205
           |
Db     203 A 203


RESULT   3
Q9Y6G5
ID   Q9Y6G5      PRELIMINARY;      PRT;   202 AA.
AC   Q9Y6G5;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   PTD002.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PITUITARY TUMOR;
RA   FU G., HUANG Q., SONG H., PENG J., ZHANG Q., MAO M., DAI M., MAO Y.,
RA   ZHOU J., CHEN Z., CHEN J.;
RT   "Human PTD002 mRNA, complete cds.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF078857; AAD44489.1; -.
SQ   SEQUENCE   202 AA;  22966 MW;  51317189 CRC32;


  Query Match             7.9%;  Score 89.5;  DB 4;  Length 202;
  Best Local Similarity   18.4%;  Pred. No. 0.83;
  Matches   38;  Conservative   44;  Mismatches   81;  Indels   43;  Gaps    6;

Qy      31 PEVAAMARLLGDLDRSTFRKLLKFVVSSL----QGEDCRDGVQRLGVSANLPEEQLGALL 86
           | :     |:  :|   | :||  ::  |  '  :     :  ::|  : :| :: |  :|
Db      13 PSMKKAVSLINAIDTGRFPRLLTRILQKLHLKAESSFSEEEEEKLQAAFSLEKQDLHLVL 72

Qy      87 AGMHTLLQQALRLPPTSLKPDTFRDQLQELCIPQDLVGDLASVVFGSQRPLLDSVAQQQG 146
             :  :|:|||:     ::||   : ||: : : ||       :           |: |:
Db      73 ETISFILEQAVY---HNVKPAALQQQLENIHLRQDKAEAFVNT--------WSSMGQET- 121

Qy     147 AWLPHVADFR------------WRVDVAISTSALARSLQPSVLMQLKLSDGSAYRFEVP 193
                |  ||              |:::: :: || |:   |  ::|| :::  :   |
Db     121 -----VEKFRQRILAPCKLETVGWQLNLQMAHSAQAKLKSPQAVLQLGVNNEDSKSLEKV 175

Qy     194 TAKFQELRYSVALVLKEMADLEKRCE 219
             :|          ||: |   : |
Db     176 LVEFSH---------KELFDFYNKLE 192


RESULT   4
Q08859
ID   Q08859      PRELIMINARY;      PRT;   413 AA.
AC   Q08859;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   FIMK PROTEIN.
GN   FIMK.
OS   Klebsiella pneumoniae.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Klebsiella.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 89155420.
RA   GERLACH G.F., CLEGG S., ALLEN B.L.;
RT   "Identification and characterization of the genes encoding the type 3
RT   and type 1 fimbrial adhesins of Klebsiella pneumoniae.";
RL   J. Bacteriol. 171:1262-1270(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   NICHOLS N.N., CLEGG S.;
RL   Submitted (AUG-1993) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; L23111; AAA25064.1; -.
KW   Fimbria.
SQ   SEQUENCE   413 AA;  46940 MW;  8149AFC3 CRC32;


  Query Match             7.8%;  Score 88.5;  DB 2;  Length 413;
  Best Local Similarity   26.8%;  Pred. No. 2.7;
  Matches   52;  Conservative   31;  Mismatches   72;  Indels   39;  Gaps   10;

Qy      52 LKFVVSSLQGEDCRDGVQRLGVSANLPEEQLGALLAGMHTLLQ----QALRLPPTSLKPD 107
           || ::: |||:   :  ||||:|      |    |||:  |::     | | | | | |
Db      94 LKVILALLQGQTIGEQAQRLGLSQKTLYTQ---RLAGVKKLVECHPHLAPRFPRT-LLPR 149
```

```
Qy    108 TFRDQLQELCIPQDLVGDLAS-VVFGSQRPLLDSVAQQQG----------------AWL 149
          :  : |      |: |  :    ||   :|::|| :| ||                  :|
Db    150 SPANAL--TAFEQEWVQAIHDRQVFPVFQPIVDSRSQLQGVEILIRWRHRGQVLHPQTFL 207

Qy    150 PHV-ADFRWRVDVAISTSALARSLQP-------SVLMQLKLSDGSAYRFEVPTAKFQELR 201
          ||  ||: | :  |       :::          || :   |:|  :    |  |: |:||
Db    208 PHFRADYTWLLLTAFVLQEAVQNINEYPGTFYFSVNIPSSLADSDSLLRMVEAAR-QQLR 266

Qy    202 Y--SVALVLKEMAD 213
              || :: | |:
Db    267 QPEGVARLVLEYAE 280


RESULT   5
O48693
ID   O48693      PRELIMINARY;      PRT;   332 AA.
AC   O48693;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   F3I6.20 PROTEIN.
GN   F3I6.20.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC   Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   FEDERSPIEL N.A., PALM C.J., CONWAY A.B., KURTZ D.B., CONWAY A.R.,
RA   AU M., ARAUJO R., BUEHLER E., DEWAR K., FENG J., KIM C., LI Y.,
RA   OJI O., OSBORNE B.I., SHINN P., SUN H., TORIUMI M., VYSOTSKAIA V.S.,
RA   YU G., ECKER J., THEOLOGIS A., DAVIS R.W.;
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC002396; AAC00587.1; -.
DR   MENDEL; 27893; Arath;3357;27893.
SQ   SEQUENCE   332 AA;  35831 MW;  A0D178A9 CRC32;


  Query Match             7.6%;  Score 86.5;  DB 10;  Length 332;
  Best Local Similarity   22.2%;  Pred. No. 3;
  Matches   51;  Conservative   27;  Mismatches   75;  Indels   77;  Gaps    8;

Qy     28 QLPPEVAAMARLLG---DLDRSTFRKLLKF----------------VVSSLQGEDCR-DG 67
          :|  |:  | | |    :: : | ::: |                    :|  : || :|
Db    112 ELSSEIDMMGRRLDANTEVIQETIQEVAKLQDGTSFIKDDVKAVFDAFENLASKVCRIEG 171

Qy     68 VQRLGVSANLPEEQLGALLAGMHTLLQQALRLPPTSLKPDTFRDQLQELCIPQDLVGDLA 127
          | |  :::||     || |       : | ||| |  ||
Db    172 NQALPSTSSLP-----ALEAAPMAPSSKTLSLPPAS--PDE------------------- 206

Qy    128 SVVFGSQRPLLDSVAQQQGAWLPHVADFRWRVDVAISTSAL-------------ARSLQP 174
              || |   :|||:    | |         |:  |:|:              |
Db    206 -----SQSPSTPNVAQKSRGLLQHTQSMSGLKDINESSSSHNTSSNGIYFGGNGASGSSS 260

Qy    175 SVLMQLKLSDGSAYRFEVPTAKFQELRYSVALVLKEMADLEKRCERRLQD 224
          || ::    |  |  |               ::|||: | |||  :  ||:
Db    261 GVLGRINKEGGMLYLVEY------------VLLKELLDNEKRLKVNLQE 297


RESULT   6
Q9XZ29
ID   Q9XZ29      PRELIMINARY;      PRT;  1212 AA.
AC   Q9XZ29;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   HYPOTHETICAL 135.8 KD PROTEIN.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   RUBIN G.M., WAN K.H., HARVEY D., LEWIS S.E., BROKSTEIN P., TSANG G.,
RA   AGBAYANI A., ARCAINA T.T., BAXTER E., BLAZEJ R.G., BUTENHOFF C.,
RA   CHAMPE M., CHAVEZ C., CHEW M., DOYLE C.M., FARFAN D.E., FRISE E.,
RA   GALLE R., GEORGE R.A., HARRIS N.L., HOSKINS R.A., EVANS-HOLM M.,
RA   HOUSTON K.A., HUMMASTI S.R., KIM E., LI P., MOSHREFI M., PACLEB J.M.,
RA   PARK S., SEQUEIRA A., SETHI H., SNIR E., SVIRSKAS R.R., WEINBURG T.,
RA   CELNIKER S.E.;
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF132186; AAD34774.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   1212 AA;  135812 MW;  518E0F63 CRC32;


  Query Match             7.6%;  Score 86.5;  DB 5;  Length 1212;
  Best Local Similarity   27.2%;  Pred. No. 17;
  Matches   56;  Conservative   30;  Mismatches   83;  Indels   37;  Gaps   10;

Qy     30 PPEVAAMARLLGDLDRSTFRKLLKFVVSSLQGEDC----------RDGVQRLGVSANLP 78
          | | |  ||:|  |:     ::    :| || : |           :||| || |  :
Db    813 PAEAAEQARILASLEEEL--EMRNAQISDLQQKVCPTDLDSRIRSLAEGVQSLGESRTVS 870

Qy     79 EEQLGALLAGMHTLLQQALRLPPTSLKPD-TFRDQLQELCIPQDLVGDLASVVFGSQRPL 137
          :: |         ||:||  ||  :||    |  |:|:  :       | ||     :  |
Db    871 KQLL-------KTLVQQR-RLQASSLNEQRTTLDELRAQLLDAQQQEDAAS----KRLRL 918

Qy    138 LDSVAQQQGAWLPHVADFRWRVDVAISTS----ALARS----LQPSVLMQLKLSDGSAYR 189
          | |  ::|   |      :  :| | | |:    | |||     | : :::  ||   | :
Db    919 LQSQHEEQ--MLAQQRAYEEKVSVLIRTANQRWAEARSPAEDQQRNQILEELLSSREALQ 976

Qy    190 FEVPTAKFQELRYSVALVLKEMADLE 215
           |:   : :    | | |  |  ||:
Db    977 QELDKLRAKNKSKSKA-VKSEPQDLD 1001


RESULT   7
O88480
ID   O88480      PRELIMINARY;      PRT;  2182 AA.
AC   O88480;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   CAIN.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98325042.
RA   LAI M.M., BURNETT P.E., WOLOSKER H., BLACKSHAW S., SNYDER S.H.;
RT   "Cain, a novel physiologic protein inhibitor of calcineurin.";
RL   J. Biol. Chem. 273:18325-18331(1998).
DR   EMBL; AF061947; AAC40176.1; -.
SQ   SEQUENCE   2182 AA;  242811 MW;  F4CC8540 CRC32;


  Query Match             7.6%;  Score 85.5;  DB 11;  Length 2182;
  Best Local Similarity   20.8%;  Pred. No. 45;
  Matches   49;  Conservative   34;  Mismatches   90;  Indels   63;  Gaps    8;

Qy     10 YLHHPGDSHSGRVSFLGAQLPPEV-----------AAMARLLGDLDRSTFRK-LLKFVV 56
          |||      :  :: :     |||:           |:: :|||  |     : |: |:
Db   1236 YLHEEAARYPKKIHY---HNPPELAMEALEVYFRLHASILKLLGKPDSGVSAEVLVSFMK 1292

Qy     57 SSLQG----------------------EDCRDGVQRL-GVSANLPEEQLGALLAGMHTL 92
          : :|                      ||      | |  |:||      | :  :|
Db   1293 EAAEGPFARGEEKNTPKASEKEKACLVDEDSHSSAGTLPGPGASLPSSSGPGLTSPPYTA 1352

Qy     93 L------------QQALRLPPTSLKPDTFRDQLQELCIPQDLVGDLASVVFGSQRPL-- 138
                      |||    |     |:    | :    ||   :  |:: ||:: |
Db   1353 TPIDHDYVKCKKPRQQA---TPDDRSQDSTAVALSDSSSTQDFFNEPTSLLDGSRKLLPE 1409
```

```
Qy    138 -----LDSVAQQQGAWLPHVADFRWRVDVAISTSALARSLQPSVLMQLKLSDGSAY 188
Db   1410 KRISGLSAQAGPSGKDLPGPTEERGKTEESLESTEAFRVVEPSVQKPVADSSASAY 1465


RESULT   8
O88031
ID   O88031      PRELIMINARY;      PRT;   218 AA.
AC   O88031;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   HYPOTHETICAL 23.7 KD PROTEIN.
GN   SC5A7.25C.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   SEEGER K.J., HARRIS D.;
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE; 97000351.
RA   REDENBACH M., KIESER H.M., DENAPAITE D., EICHNER A., CULLUM J.,
RA   KINASHI H., HOPWOOD D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
DR   EMBL; AL031107; CAA19954.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   218 AA;  23670 MW;  19769615 CRC32;


  Query Match             7.3%;  Score 83;  DB 2;  Length 218;
  Best Local Similarity   27.1%;  Pred. No. 3.6;
  Matches   48;  Conservative   22;  Mismatches   51;  Indels   56;  Gaps    9;

Qy     68 VQRLGVSANLPEEQLGALLAGMHT--------LLQQALRLPPTSLKPDTFRDQLQELCIP 119
Db     40 VGRLSAASGVPESVVGALLGGRPTGGADVQTRFVQRLDLLRRTRLKPNGRKYTQQEIA-- 98

Qy    120 QDLVG----DLASVVFGSQRPLLDSVAQQQGAWLPHVADFRWRVDVAISTSALARSLQPS 175
Db     98 -DGAGMSRQQAGALINGDRRPTMEHCDAIQRFFRVH-AGFL----TAEDPEALAGALQ-- 150

Qy    176 VLMQLKLSDGSAYRFEVPTAKFQELRYSVALVLKEMADLEKRC--------ERRLQD 224
Db    150 ------------RTE------QEL-------LQQLADRERQAAAPADDPLERLLQD 180


RESULT   9
Q9Y2F1
ID   Q9Y2F1      PRELIMINARY;      PRT;   418 AA.
AC   Q9Y2F1;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   KIAA0942 PROTEIN (FRAGMENT).
GN   KIAA0942.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 99246063.
RA   NAGASE T., ISHIKAWA K., SUYAMA M., KIKUNO R., HIROSAWA M.,
RA   MIYAJIMA N., TANAKA A., KOTANI H., NOMURA N., OHARA O.;
RT   "Prediction of the coding sequences of unidentified human genes. XIII.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 6:63-70(1999).
DR   EMBL; AB023159; BAA76786.1; -.
FT   NON_TER      1      1
SQ   SEQUENCE   418 AA;  47378 MW;  02662B38 CRC32;


  Query Match             7.3%;  Score 83;  DB 4;  Length 418;
  Best Local Similarity   23.8%;  Pred. No. 8.6;
  Matches   57;  Conservative   41;  Mismatches   98;  Indels   44;  Gaps   13;

Qy      1 MSAVGAAT-PYLHHPGDSHSG--RVSFLGAQLPPEVAAMARLLGDLDRSTFRKLLKFVVS 57
Db    135 ISRIGSTTNPFLDIPHDPNAAVYKSGFLARKIHADMDGKKTPRGKRGWKTFYAVLKGTVL 194

Qy     58 SLQGEDCRDGVQRLGVSANLPEEQLGALLAGMHTLLQQAL---RLPPT-SLKPDTFRDQL 113
Db    195 YLQKDEYKP-------EKALSEEDLKNAVSVHHALASKATDYEKKPNVFKLKTADWRVLL 247

Qy    114 QELCIPQDLVG-----DLASVVF---------GSQ----RPLLDSVA---QQQGAWLPHV 152
Db    248 FQTQSPEEMQGWINKINCVAAVFSAPPFPAAIGSQKKFSRPLLPATTTKLSQEEQLKSHE 307

Qy    153 ADFRWRVDVAISTS-ALARSLQPSVLMQLKLSDGSAYRFEVPTAKFQELRYSVAL-VLKE 210
Db    308 SKLK-----QITTELAEHRSHPPD--KKVKAKDVDEYKLKDHYLEFEKTRYEMYVSILKE 360


RESULT  10
Q9WU50
ID   Q9WU50      PRELIMINARY;      PRT;   992 AA.
AC   Q9WU50;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   DYSTONIN (FRAGMENT).
GN   BPAG1-N.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C;
RX   MEDLINE; 99268151.
RA   LEUNG C.L., SUN D., LIEM R.K.H.;
RT   "The intermediate filament protein peripherin is the specific
RT   interaction partner of mouse BPAG1-n (dystonin) in neurons.";
RL   J. Cell Biol. 144:435-446(1999).
DR   EMBL; AF115383; AAD22959.1; -.
FT   NON_TER      1      1
SQ   SEQUENCE   992 AA;  112950 MW;  A0959F08 CRC32;


  Query Match             7.2%;  Score 82;  DB 11;  Length 992;
  Best Local Similarity   22.4%;  Pred. No. 33;
  Matches   59;  Conservative   30;  Mismatches  106;  Indels   68;  Gaps   11;

Qy      2 SAVGAATPYLHHPGDSHSGRVSFL-------------------------GAQLPPEVAA 35
Db    223 SAFDTAGRECHHPAEISPGNSGHLNLKTRLPLSRWTQEPHQTEGKWPHRAAEQLPKEV-- 281

Qy     36 MARLLG-DLDRSTFRKLLKFVVSSLQGE-----DCRDGVQRLGVSANLPEEQLGALLAGM 89
Db    281 QFRQPGAPLDRESSQPCYSEYFSQTSTELQITFDDKNPITR-------------LSEL 325

Qy     90 HTLLQQALR--LPPTSLKPDTFRDQLQELCIPQDLV-----GDLASVVFGSQRPLLDSVA 142
Db    326 ETMREQALHPSRPPVTYQDDKLERELVKLLTPLEIAKNKQCGMHTEVTTLKQEKRLGSSA 385
```

```
Qy    143 QQQGAWLPHVADFRWRVDVAISTSALARSLQPSVLMQLKLSDGSAYRFEVPTAKFQELRY 202
          | |:       |    :    | :|::|     | |:   : | |    :|| ||:
Db    386 ---GGWMLEGC----RTSGGLKGDFLKKSVEPEASPSLDLNQACSVRDE--EFQFQGLRH 436

Qy    203 SVA---LVLKEMADLEKRCERRL 222
          :|    ||  :: |:    : ||
Db    437 TVTGRQLVEAKLLDMRTVEQLRL 459


RESULT  11
P95629
ID   P95629      PRELIMINARY;      PRT;  1224 AA.
AC   P95629;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   PUTA GENE.
GN   PUTA.
OS   Rhizobium meliloti.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Sinorhizobium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GR4;
RX   MEDLINE; 97157944.
RA   JIMENEZ-ZURDO J.I., GARCIA-RODRIGUEZ F.M., TORO N.;
RT   "The Rhizobium meliloti putA gene: its role in the establishment of
RT   the symbiotic interaction with alfalfa.";
RL   Mol. Microbiol. 23:85-93(1997).
DR   EMBL; Y08500; CAA69727.1; -.
DR   PFAM; PF00171; aldedh; 1.
DR   PFAM; PF01619; Pro_dh; 1.
SQ   SEQUENCE   1224 AA;  131190 MW;  B2C95AE0 CRC32;


  Query Match             7.2%;  Score 82;  DB 2;  Length 1224;
  Best Local Similarity   28.1%;  Pred. No. 44;
  Matches   25;  Conservative   21;  Mismatches   33;  Indels   10;  Gaps    3;

Qy     29 LPP--EVAAMARLLGDLDRSTFRKLLKFVVSSLQGEDCRDGVQRLGVSANLPEEQLGALL 86
          |||  | |  :: : |   || |||    : :|:|:    ||: |    :|  ::  ||:
Db     47 LPPLVEAATQSKEIRDAAASTARKL----IEALRGKHSGSGVEGLVQEYSLSSQEGVALM 102

Qy     87 AGMHTLLQQALRLPPTSLKPDTFRDQLQE 115
             | :: :|:| |: :    ||:: :
Db    103 C----LAERPVRIPDTATRDALIRDKIAD 127


RESULT  12
Q62037
ID   Q62037      PRELIMINARY;      PRT;   800 AA.
AC   Q62037;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   ROD PHOSPHODIESTERASE BETA SUBUNIT ISOZYME.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=RETINA;
RX   MEDLINE; 91130581.
RA   BAEHR W., CHAMPAGNE M.S., LEE A.K., PITTLER S.J.;
RT   "Complete cDNA sequences of mouse rod photoreceptor cGMP
RT   phosphodiesterase alpha- and beta-subunits, and identification of
RT   beta'-, a putative beta-subunit isozyme produced by alternative
RT   splicing of the beta-subunit gene.";
RL   FEBS Lett. 278:107-114(1991).
DR   EMBL; X87952; CAA61202.1; -.
DR   PROSITE; PS00126; PDEASE_I; 1.
DR   PFAM; PF01590; GAF; 2.
DR   PFAM; PF00233; PDEase; 1.
DR   PRINTS; PR00387; PDIESTERASE1.
SQ   SEQUENCE   800 AA;  92076 MW;  3204E9D4 CRC32;


  Query Match             7.2%;  Score 81.5;  DB 11;  Length 800;
  Best Local Similarity   19.4%;  Pred. No. 28;
  Matches   49;  Conservative   38;  Mismatches   68;  Indels   97;  Gaps   10;

Qy     32 EVAAMARLLGDLDRS-----TFRKLLKFVVSSLQGEDC-------RDGVQRLGVSANLPE 79
          | ||:  |: |:  |        |:|: : : |  : |       |:|:  |
Db     55 ESAALFELVQDMQESVNMERVVFKILRRLCTILHADRCSLFMYRQRNGIAEL-------- 107

Qy     80 EQLGALLAGMHTLLQQALRLPPTSLKPDTFRDQLQELCIPQDLVGDLASVVFGSQRPLLD 139
                    | ||   |::||:   |::  :| |     : :||     ::
Db    107 ----------------ATRL--FSVQPDSL---LEDCLVPPD-----SEIVFPLDIGIVG 140

Qy    140 SVAQ-------QQGAWLPHVADFRWRVDVAISTSALARSLQ-----PSVLMQLKLSDGS- 187
          |||       |  | || : |   :   :: : |:  :       :|:| :   ||
Db    141 HVAQTKKMINVQDVAECPHFSSFADELTDYVTKNILSTPIMNGKDVVAVIMAVNKLDGPC 200

Qy    187 ------------------------AYRFEVPTAKFQELRYSVALVLKEMADLEKR---- 218
                                   :|     | : | | :|   | :|: |:|::
Db    201 FTSEDEDVFTKYLNFATLNLKIYHLSYLHNCETRRGQVLLWSANKVFEELTDIERQFHKA 260

Qy    218 ---------CER 220
                   |||
Db    261 FYTVRAYLNCER 272


RESULT  13
O28746
ID   O28746      PRELIMINARY;      PRT;   145 AA.
AC   O28746;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   CONSERVED HYPOTHETICAL PROTEIN.
GN   AF1526.
OS   Archaeoglobus fulgidus.
OC   Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC   Archaeoglobus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX   MEDLINE; 98049343.
RA   KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
RA   KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,
RA   RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRPIDES N.C.,
RA   FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
RA   KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
RA   PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,
RA   OVERBEEK R., GOCAYNE J.D., WEIDMAN J.F., MCDONALD L., UTTERBACK T.,
RA   COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M.,
RA   SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
RA   MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,
RA   VENTER J.C.;
RT   "The complete genome sequence of the hyperthermophilic, sulphate-
RT   reducing archaeon Archaeoglobus fulgidus.";
RL   Nature 390:364-370(1997).
DR   EMBL; AE000997; AAB89722.1; -.
DR   HSSP; Q57997; 1MJH.
DR   TIGR; AF1526; -.
KW   Hypothetical protein.
SQ   SEQUENCE   145 AA;  16339 MW;  E5041F32 CRC32;


  Query Match             7.2%;  Score 81;  DB 1;  Length 145;
  Best Local Similarity   30.7%;  Pred. No. 3.2;
  Matches   35;  Conservative   21;  Mismatches   40;  Indels   18;  Gaps    5;

Qy     36 MARLLGDLDRS-----TFRKLLKFVVSSLQGEDCRDGVQRLGVSANLPEEQLGALLAGMH 90
          | |:|  || :      |:|  |     |:||     :: : |   :|||:    ||  |
```

```
Db      1 MERILLVLDDTGRGEIAFQKLKKLAEDGLRGEVYILYIREMEVPPFVPEEK---ELAAYH 57

Qy     91 TLLQQALRLPPTSLKPDTFRDQLQELCIPQDLVGDLASVVFGSQRPLLDSVAQQ 144
          |: |:::      | : |::||::  :    | |: |||||     |  | :|
Db     58 RLMTQSMK------KLEGFKNQLEKAGLK---VSDV-SVVFGKYADRLLLVEKQ 101


RESULT  14
Q9WZJ7
ID   Q9WZJ7      PRELIMINARY;      PRT;   161 AA.
AC   Q9WZJ7;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   HYPOTHETICAL 18.1 KD PROTEIN.
GN   TM0738.
OS   Thermotoga maritima.
OC   Bacteria; Thermotogales; Thermotoga.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 99287316.
RA   NELSON K.E., CLAYTON R.A., GILL S.R., GWINN M.L., DODSON R.J.,
RA   HAFT D.H., HICKEY E.K., PETERSON J.D., NELSON W.C., KETCHUM K.A.,
RA   MCDONALD L., UTTERBACK T.R., MALEK J.A., LINHER K.D., GARRETT M.M.,
RA   STEWART A.M., COTTON M.D., PRATT M.S., PHILLIPS C.A., RICHARDSON D.,
RA   HEIDELBERG J., SUTTON G.G., FLEISCHMANN R.D., WHITE O., SALZBERG S.L.,
RA   SMITH H.O., VENTER J.C., FRASER C.M.;
RT   "Evidence for lateral gene transfer between Archaea and bacteria from
RT   genome sequence of Thermotoga maritima.";
RL   Nature 399:323-329(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   NELSON K.E., CLAYTON R.A., GILL S.R., GWINN M.L., DODSON R.J.,
RA   HAFT D.H., HICKEY E.K., PETERSON J.D., NELSON W.C., KETCHUM K.A.,
RA   MCDONALD L., UTTERBACK T.R., MALEK J.A., LINHER K.D., GARRETT M.M.,
RA   STEWART A.M., COTTON M.D., PRATT M.S., PHILLIPS C.A., RICHARDSON D.,
RA   HEIDELBERG J., SUTTON G.G., FLEISCHMANN R.D., WHITE O., SALZBERG S.L.,
RA   SMITH H.O., VENTER J.C., FRASER C.M.;
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE001744; AAD35819.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   161 AA;  18092 MW;  B3E9F115 CRC32;


  Query Match             7.1%;  Score 80.5;  DB 2;  Length 161;
  Best Local Similarity   27.7%;  Pred. No. 4.1;
  Matches   26;  Conservative   20;  Mismatches   35;  Indels   13;  Gaps    3;

Qy      1 MSAVGAATP-YLHHPGDSHSGRVSFLGAQLPPEVAAMARLLGDLDRSTFRKLLKFVVSSL 59
          ::||   |  ::: ||   | |   ||::|  :   : | : :|:|      ||     |
Db     69 LAAVAGRTQVFIYTPGTKSSERTHSLGSELQDKQQEIERKILELERQ-----LKTETDPL 123

Qy     60 QGEDCRDGVQRLGVSANL-------PEEQLGALL 86
          : |  |: ::|| :: |:       ||  :| ||
Db    124 ERERLREDLERLKLALNILKASLRAPELLVGVLL 157


RESULT  15
Q9ZGG9
ID   Q9ZGG9      PRELIMINARY;      PRT;   311 AA.
AC   Q9ZGG9;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   EXOPOLYPHOSPHATASE PPX (FRAGMENT).
GN   PPX.
OS   Heliobacillus mobilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Heliobacterium group; Heliobacillus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 99061957.
RA   XIONG J., INOUE K., BAUER C.E.;
RT   "Tracking molecular evolution of photosynthesis by characterization of
RT   a major photosynthesis gene cluster from Heliobacillus mobilis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:14851-14856(1998).
DR   EMBL; AF080002; AAC84010.1; -.
FT   NON_TER       1       1
SQ   SEQUENCE   311 AA;  33883 MW;  B91E34DE CRC32;


  Query Match             7.1%;  Score 80;  DB 2;  Length 311;
  Best Local Similarity   22.0%;  Pred. No. 11;
  Matches   59;  Conservative   32;  Mismatches   81;  Indels   96;  Gaps   14;

Qy      5 GAATPYLHHPGDSHSGRVSFLGAQLPPEVAAMARLLGDLDRS---------TFRKLLKFV 55
            |  ||  ||   :  ||       | |:  |: | :  |           : :::  |
Db     27 GKQTPLWHHLDTTRLGREVTRSGTLHPD--AVERTMAALGHGVESMQKYGISVQQVAAFA 84

Qy     56 VS----SLQGED-CRDGVQRLGVSANLPEEQLGALLAGM---HTLLQQALRLPPTSLKPD 107
           |    :| |:| ||   :|  :           |||:|    |   | |:   |  |
Db     85 TSAVREALNGQDFCRRVYERFRLEV--------ALLSGETEGHLSYQGAVEGLPEGLAAG 136

Qy    108 TFRDQLQELCIPQDLVGDLASVVFGSQRPLLDSVAQQQGAWLPHVADFRW---------- 158
                : :  |: |  | |||                |  |  |
Db    137 -------RIPVVVDIGGGSAEVVF-----------------HHKDRWWRQSFPLGAVR 170

Qy    158 -------RVDVAI-------STSALARSLQP------------SVLMQLK-LSDGSAYR 189
                |||:|          |:| |  :|             :: :||:  :    :
Db    171 LTESPLGRVDIASVWAPAVEKVSSLQRMGRPILIGVGGTITTVAAIALQLRDYAPDKVHG 230

Qy    190 FEVPTAKFQELRYSVALVLKEMADLEKR 217
          :|:|  | :    |:|  |: | |||:|
Db    231 YEIPLEKVK----SIARELRAM-DLEER 253


Search completed: May 15, 2000, 05:33:29
Job time: 12455 sec
```

OM protein - protein search, using sw model

Run on: May 14, 2000, 23:18:10 ; Search time 38.86 Seconds
(without alignments)
136.533 Million cell updates/sec

Title: US-09-223-796-2
Perfect score: 1112
Sequence: 1 MSALGAAAPYLHHPADSHSG.....ALVLKEMAELEKKCERKLQD 224

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : A_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1112	100.0	224	1	W37723	Rat Hypertension r
2	99.5	8.9	359	1	W29380	S. pneumoniae pept
3	99.5	8.9	359	1	W38592	S. pneumoniae pept
4	86.5	7.8	195	1	W71684	Amino acid sequenc
5	86.5	7.8	196	1	Y07867	Human secreted pro
6	84.5	7.6	1786	1	W24790	P. falciparum live
7	82.5	7.4	289	1	W53924	Pseudomonas pseudo
8	82.5	7.4	758	1	W46270	Moraxella catarrha
9	82.5	7.4	1196	1	R75189	Osteoinductive ret
10	82	7.4	427	1	P90476	Polypeptide of hum
11	82	7.4	536	1	R76502	Human MIS mature p
12	82	7.4	559	1	P70196	Sequence encoded b
13	82	7.4	560	1	R76501	Human MIS protein.
14	81	7.3	3588	1	R34712	Bacillus subtilis
15	80.5	7.2	289	1	R88023	Mature Pseudomonas
16	80	7.2	1048	1	W97720	Staphylococcus aur
17	79.5	7.1	289	1	R88021	Mature Pseudomonas
18	79.5	7.1	875	1	W48309	Pisum sativum ACCa
19	79	7.1	660	1	Y13349	Amino acid sequenc
20	79	7.1	1145	1	W32097	Miniature swine re
21	79	7.1	1194	1	W39272	Porcine retrovirus
22	78.5	7.1	169	1	W88390	Human Zneul partia
23	78.5	7.1	181	1	W88391	Human Zneul partia
24	78.5	7.1	254	1	W88382	Human neuro-growth
25	78.5	7.1	273	1	W88381	Human neuro-growth
26	78.5	7.1	363	1	W06320	Human mitogen-acti
27	78.5	7.1	363	1	W97670	Human mitogen acti
28	78.5	7.1	393	1	W06321	Human mitogen-acti
29	78.5	7.1	393	1	W97671	Human mitogen acti
30	78.5	7.1	399	1	W06322	Human mitogen-acti
31	78.5	7.1	399	1	W97672	Human mitogen acti
32	78.5	7.1	1427	1	R10534	Human 160kD mediat
33	78	7.0	687	1	W41586	Truncated restin p
34	77.5	7.0	289	1	R88025	Mature Pseudomonas
35	77.5	7.0	289	1	R88024	Mature Pseudomonas
36	77	6.9	3567	1	R44431	eryA region polype
37	76.5	6.9	289	1	R88022	Mature Pseudomonas
38	76.5	6.9	403	1	W72844	Polymerase enhanci
39	76.5	6.9	1078	1	R28963	Notch hN3k full le
40	76.5	6.9	1527	1	W81172	Human BAZ1-beta pr
41	76	6.8	388	1	W18029	Huntingtin interac
42	76	6.8	914	1	W18030	Huntingtin interac
43	75.5	6.8	477	1	P50636	Angiotensin. New c
44	75	6.7	344	1	W32425	Mycobacterium tube
45	75	6.7	344	1	W32357	Mycobacterium tube

ALIGNMENTS

RESULT 1
W37723
ID W37723 standard; Protein; 224 AA.
AC W37723;
DT 09-JUN-1998 (first entry)
DE Rat Hypertension related calcium regulator.
KW Hypertension related calcium regulated gene; HCaRG; rat parathyroid;
KW extracellular calcium concentration; antibody; hypertension;
KW hyperthyroidism; osteoporosis; heart failure; diabetes; stroke;
KW cancer; inflammatory disease; asthma.
OS Rattus rattus.
FH Key Location/Qualifiers
FT Misc_difference 15. .21
FT /note= "EF-hand like motif"
PN WO9749807-A2.
PD 31-DEC-1997.
PF 23-JUN-1997; CA0439.
PR 21-JUN-1996; US-667495.
PA (GOSS/) GOSSARD F.
PA (HAME/) HAMET P.
PA (LEWA/) LEWANCZUK R.
PA (TREM/) TREMBLAY J.
PI Gossard F, Hamet P, Lewanczuk R, Tremblay J;
DR WPI; 98-077171/07.
DR N-PSDB; V18890.
PT Hypertension related calcium regulated gene - useful to develop
PT products to treat or detect, e.g. hypertension, stroke,
PT osteoporosis, heart failure, cancer, diabetes or asthma
PS Claim 8; Pages 26-27; 46pp; English.
CC This is the amino acid sequence of the hypertension related calcium
CC regulated gene (HCaRG), which was isolated from the rat parathyroid.
CC Its expression is regulated by extracellular calcium concentration.
CC An antibody against the protein, can be used to detect or modulate
CC (e.g. enhance or inhibit) abnormal calcium levels. They can
CC specifically be used to detect or treat, e.g. hypertension,
CC hyperthyroidism, osteoporosis, heart failure, diabetes, stroke, cancer,
CC inflammatory disease, and asthma.
SQ Sequence 224 AA;

Query Match 100.0%; Score 1112; DB 1; Length 224;
Best Local Similarity 100.0%; Pred. No. 1.1e-101;
Matches 224; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MSALGAAAPYLHHPADSHSGRVSFLGSQPSPEVTAVAQLLKDLDRSTFRKLLKLVVGALH 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MSALGAAAPYLHHPADSHSGRVSFLGSQPSPEVTAVAQLLKDLDRSTFRKLLKLVVGALH 60

Qy      61 GKDCREAVEQLGASANLSEERLAVLLAGTHTLLQQALRLPPASLKPDAFQEELQELGIPQ 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GKDCREAVEQLGASANLSEERLAVLLAGTHTLLQQALRLPPASLKPDAFQEELQELGIPQ 120

Qy     121 DLIGDLASLAFGSQRPLLDSVAQQQGSSLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQL 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 DLIGDLASLAFGSQRPLLDSVAQQQGSSLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQL 180

Qy     181 KLTDGSAHRFEVPIAKFQELRYSVALVLKEMAELEKKCERKLQD 224
```

```
Db    181 KLTDGSAHRFEVPIAKFQELRYSVALVLKEMAELEKKCERKLQD 224
```

RESULT 2
W29380
ID W29380 standard; Protein; 359 AA.
AC W29380;
DT 20-MAY-1998 (first entry)
DE S. pneumoniae peptide releasing factor RF-1.
KW RF-1; peptide releasing factor; Streptococcus pneumoniae; diagnosis;
KW otitis media; conjunctivitis; meningitis; pneumonia; endocarditis;
KW bacterial infection.
OS Streptococcus pneumoniae.
PN WO9743304-A1.
PD 20-NOV-1997.
PF 14-MAY-1997; U08272.
PR 14-MAY-1996; US-017670.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PA (SMIK) SMITHKLINE BEECHAM PLC.
PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO,
PI Stodola RK;
DR WPI; 98-008794/01.
DR N-PSDB; T89027.
PT DNA encoding peptide releasing factor, RF-1 from Streptococcus
PT pneumoniae - useful for diagnosis and treatment of, e.g. otitis
PT media, conjunctivitis and meningitis, etc.
PS Claim 12; Page 9; 41pp; English.
CC This is a peptide releasing factor (prfA) RF-1 from Streptococcus
CC pneumoniae. RF-1 is one of the three factors required for the
CC termination of protein synthesis in bacteria. Upon recognition of stop
CC codons, RF-1 binds to the ribosome resulting in the peptidyl transferase
CC centre of the ribosome switching to a hydrolytic function to remove the
CC completed polypeptide chain. A mutation in S. typhimurium prfA has been
CC demonstrated to inhibit cell division through a novel regulatory circuit.
CC Hence, RF-1 may be used as an anti-bacterial target. RF-1 can be used to
CC identify compounds which interact with and inhibit or activate its
CC activity. RF-1 antagonists can be used to treat diseases caused by
CC S. pneumoniae. Through genetic immunisation, immunological response in a
CC mammal can be induced by inoculation with RF-1 or delivery of RF-1
CC encoding nucleic acid in a vector, adequate to produce antibody and/or
CC T cell immune responses to protect the animal from disease. The products
CC and methods are particularly useful for diagnosis of disease, preferably,
CC bacterial infections caused by S. pneumoniae, especially otitis media,
CC conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural
CC empyema and endocarditis.
SQ Sequence 359 AA;

Query Match 8.9%; Score 99.5; DB 1; Length 359;
Best Local Similarity 24.6%; Pred. No. 0.054;
Matches 49; Conservative 37; Mismatches 58; Indels 55; Gaps 11;

```
Qy      48 FRKLLKLVVGALHGKDCREAVEQLGASANLSEERLAVLLAGTHTLLQQALRLPPASLKPD 107
           ::::|: :|      |  | ::: |  |:| |           | :| |:   |  : :
Db      53 YKQVLQNIV------DAEEMIKESGGDADLEE-----------LAKQELK--DAKAEKE 92

Qy     108 AFQEELQELGIPQD------LI--------GDLASLAFGSQRPLLDSVAQQQG------- 147
           ::|:|: | :|:|      :|          || |:|  |    :    |: ||
Db      93 EYEEKLKILLLPKDPNDDKNIILEIRGAAGGDEAALFAGDLLTMYQKYAEAQGWRFEVME 152

Qy     147 SSLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQLKLTDGSAHRFE-VPIAKFQ---ELRY 202
           :|:  |  |:   :|    | |      ||  :||   | ||| : ||: : |
Db     153 ASMNGVGGFK---EVVAMVSGQ------SVYSKLKYESG-AHRVQRVPVTESQGRVHTST 202

Qy     203 SVALVLKEMAELEKKCERK 221
           :  ||: |: |:|   : |
Db     203 ATVLVMPEVEEVEYDIDPK 221
```

RESULT 3
W38592
ID W38592 standard; Protein; 359 AA.
AC W38592;
DT 09-NOV-1998 (first entry)
DE S. pneumoniae peptide chain release factor 1.
KW Streptococcus pneumoniae protein; genetic immunisation; antagonist;
KW immunological response; inoculation; antibody production; inhibitor;
KW T cell immune response; antimicrobial compound; bacterial adhesion;
KW extracellular matrix protein; protein-mediated cell invasion; wound;
KW pathogenesis.
OS Streptococcus pneumoniae.
PN WO9743303-A1.
PD 20-NOV-1997.
PF 14-MAY-1997; U07950.
PR 14-MAY-1996; US-017670.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PA (SMIK) SMITHKLINE BEECHAM PLC.
PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO,
PI Stodola RK;
DR WPI; 98-008793/01.
DR N-PSDB; T98642.
PT Novel Streptococcus pneumoniae proteins and related DNA - useful for
PT diagnosing anti-microbial agents for treatment of bacterial
PT infections
PS Claim 12; Pages 359-360; 483pp; English.
CC This sequence represents a Streptococcus pneumoniae protein that, based
CC on homology with a Bacillus subtilis protein, is a peptide chain release
CC factor 1, and is encoded by a DNA sequence of the invention.
CC The DNA sequences were isolated from Streptococcus pneumoniae strain
CC 0100993 (NCIMB 40794). The Streptococcus pneumoniae proteins of the
CC invention can be used to identify compounds which interact with and
CC inhibit or activate the activity of the proteins. Antagonists can be
CC used to treat diseases caused by S. pneumoniae proteins, through genetic
CC immunisation. They can also be used to induce an immunological response
CC in a mammal by inoculation with the S. pneumoniae proteins or delivery
CC of the encoding nucleic acids in a vector adequate to produce antibody
CC and/or T cell immune responses to protect the animal from disease. The
CC proteins can also be used to identify antimicrobial compounds which are
CC capable of inhibiting their bioactivity. In particular the proteins of
CC the invention can be used to prevent adhesion of bacteria to mammalian
CC extracellular matrix proteins on in-dwelling devices or in wounds, to
CC block protein-mediated mammalian cell invasion, and to block the normal
CC progression of pathogenesis in infections initiated other than by the
CC implantation of in-dwelling devices or other surgical techniques.
SQ Sequence 359 AA;

Query Match 8.9%; Score 99.5; DB 1; Length 359;
Best Local Similarity 24.6%; Pred. No. 0.054;
Matches 49; Conservative 37; Mismatches 58; Indels 55; Gaps 11;

```
Qy      48 FRKLLKLVVGALHGKDCREAVEQLGASANLSEERLAVLLAGTHTLLQQALRLPPASLKPD 107
           ::::|: :|      |  | ::: |  |:| |           | :| |:   |  : :
Db      53 YKQVLQNIV------DAEEMIKESGGDADLEE-----------LAKQELK--DAKAEKE 92

Qy     108 AFQEELQELGIPQD------LI--------GDLASLAFGSQRPLLDSVAQQQG------- 147
           ::|:|: | :|:|      :|          || |:|  |    :    |: ||
Db      93 EYEEKLKILLLPKDPNDDKNIILEIRGAAGGDEAALFAGDLLTMYQKYAEAQGWRFEVME 152

Qy     147 SSLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQLKLTDGSAHRFE-VPIAKFQ---ELRY 202
           :|:  |  |:   :|    | |      ||  :||   | ||| : ||: : |
Db     153 ASMNGVGGFK---EVVAMVSGQ------SVYSKLKYESG-AHRVQRVPVTESQGRVHTST 202

Qy     203 SVALVLKEMAELEKKCERK 221
           :  ||: |: |:|   : |
Db     203 ATVLVMPEVEEVEYDIDPK 221
```

RESULT 4
W71684
ID W71684 standard; Protein; 195 AA.
AC W71684;
DT 04-DEC-1998 (first entry)
DE Amino acid sequence of the human tumourigenesis associated protein.
KW Human; tumourigenesis associated protein; HTAP; transplantation;

KW tumour; Antagonist; cancer; inflammation; immunological disease;
KW antibody; probe; primer; PCR; amplification; hybridisation;
KW inhibition.
OS Homo sapiens.
PN WO9841635-A1.
PD 24-SEP-1998.
PF 20-MAR-1998; U06066.
PR 20-MAR-1997; US-822260.
PA (INCY-) INCYTE PHARM INC.
PI Goli SK, Hillman JL;
DR WPI; 98-521224/44.
DR N-PSDB; V58281.
PT New tumorigenesis-associated protein and related nucleic acid,
PT vectors, transformed cells - antibodies, agonists and antagonists,
PT for diagnosis, treatment and prevention of abnormal cellular
PT differentiation, particularly cancers and inflammation
PS Claim 1; Fig 1A-1B; 54pp; English.
CC This is the amino acid sequence of the human tumourigenesis
CC associated protein (HTAP), used in the method of the invention. HTAP,
CC is involved with cell proliferation and inflammation. It can be used
CC to stimulate cell proliferation (e.g. of cells intended for
CC transplantation in treatment of tumours or infections, or to treat
CC genetic defects). Antagonists of HTAP are used to treat or prevent
CC a wide range of cancers (adenocarcinoma, melanoma, sarcoma, lymphoma,
CC leukaemia etc.), also inflammation where associated with infection or
CC immunological disease (e.g. asthma, cystic fibrosis, rheumatoid
CC arthritis). HTAP is also used to raise antibodies and to screen
CC libraries for specific-binding agents. The antibodies are used as for
CC diagnosis or monitoring of HTAP-related diseases (in usual
CC immunoassays), in competitive drug screens and to isolate HTAP from
CC its natural sources. HTAP derived probes or primers, are used in
CC standard amplification or hybridisation tests to diagnose HTAP-related
CC diseases; to identify related sequences; for genomic mapping and for
CC screening for specific inhibitors.
SQ Sequence 195 AA;

Query Match 7.8%; Score 86.5; DB 1; Length 195;
Best Local Similarity 22.1%; Pred. No. 0.44;
Matches 46; Conservative 33; Mismatches 82; Indels 47; Gaps 8;

```
Qy      38 QLLKD---LDRSTFRKLLKLVVGALHGKDCREAV--------------EQLGASANLSEE 80
           |:| |    | : |  ||:    :|      |||             :   |:|
Db      12 QMLADPRSFDSNAFTLLLRAAFQSLLDAQADEAVLDHPDLKHIDPVVLKHCHAAA----- 67

Qy      81 RLAVLLAGTHTLLQQALRLPPASLKPDAFQEELQELGIPQDLIGDLASLAFGSQRPLLDS 140
             :| || |   :  |      |:   |  |  ||   :           : : :  |:
Db      67 ATYILEAGKHRADKSTL---STYLEDCKFDRERIELFCTE----------YQNNKNSLEI 113

Qy     141 VAQQQGSSLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQLKL--TDGSAH---RFEVPIA 195
           :    | ||||::   ||::  | |:   |  :|: |: | :  ||  ::    |   :
Db     114 LLGSIGRSLPHITDVSWRLEYQIKTNQLHRMYRPAYLVTLSVQNTDSPSYPEISFSCSME 173

Qy     196 KFQELRYSVALVLKEMAELEKKCERKLQ 223
           : |:|      : :: :  |  ||  |
Db     174 QLQDL-------VGKLKDASKSLERATQ 194
```

RESULT 5
Y07867
ID Y07867 standard; Protein; 196 AA.
AC Y07867;
DT 06-JUL-1999 (first entry)
DE Human secreted protein fragment encoded from gene 16.
KW Human; secreted protein; treatment; prevention; protein therapy; AIDS;
KW gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
KW developmental abnormality; fetal deficiency; blood disorder; leukemia;
KW immune system disease; autoimmune disease; hepatic disease; lymphoma;
KW renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;
KW cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;
KW pulmonary disorder; transplant rejection; osteoclast; osteoporosis;
KW arthritis; malignancy; digestive; endocrine; infection.
OS Homo sapiens.

PN WO9918208-A1.
PD 15-APR-1999.
PF 01-OCT-1998; U20775.
PR 02-OCT-1997; US-060884.
PR 02-OCT-1997; US-060833.
PR 02-OCT-1997; US-060836.
PR 02-OCT-1997; US-060837.
PR 02-OCT-1997; US-060838.
PR 02-OCT-1997; US-060839.
PR 02-OCT-1997; US-060843.
PR 02-OCT-1997; US-060862.
PR 02-OCT-1997; US-060866.
PR 02-OCT-1997; US-060874.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Carter KC, Duan DR, Endress GA, Feng P, Ferrie AM,
PI Florence KA, Greene JM, Janat F, Lafleur DW, Ni J,
PI Rosen CA, Ruben SM, Shi Y, Young P, Yu G;
DR WPI; 99-264022/22.
DR N-PSDB; X37466.
PT New isolated human genes and the secreted polypeptides they encode
PS Claim 1b; Page 285-286; 368pp; English.
CC This invention describes novel isolated human genes and the secreted
CC proteins they encode. The products of the invention are useful for
CC preventing, treating or ameliorating medical conditions, e.g. by protein
CC or gene therapy. Also pathological conditions can be diagnosed by
CC determining the amount of the new polypeptides in a sample or by
CC determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the 101 polynucleotides, based on
CC which tissues they are most highly expressed in, and include developing
CC products for the diagnosis or treatment of cancer, tumours,
CC neurodegenerative disorders, developmental abnormalities and fetal
CC deficiencies, blood disorders, leukemias, diseases of the immune system,
CC autoimmune diseases, hepatic and renal disease, lymphomas, inflammation,
CC allergies, Alzheimer's and cognitive disorders, schizophrenia, prostate
CC disease, skeletal or cardiac muscle disorders, pulmonary disorders,
CC transplant rejection, disorders involving osteoclasts such as
CC osteoporosis, arthritis or malignancies, digestive/endocrine disorders,
CC infections and AIDS. The human secreted proteins of the invention are
CC represented in Y07852-Y07993 and the encoding nucleic acids are
CC represented in X37451-X37552.
SQ Sequence 196 AA;

Query Match 7.8%; Score 86.5; DB 1; Length 196;
Best Local Similarity 22.1%; Pred. No. 0.45;
Matches 46; Conservative 33; Mismatches 82; Indels 47; Gaps 8;

```
Qy      38 QLLKD---LDRSTFRKLLKLVVGALHGKDCREAV--------------EQLGASANLSEE 80
           |:| |    | : |  ||:    :|      |||             :   |:|
Db      12 QMLADPRSFDSNAFTLLLRAAFQSLLDAQADEAVLDHPDLKHIDPVVLKHCHAAA----- 67

Qy      81 RLAVLLAGTHTLLQQALRLPPASLKPDAFQEELQELGIPQDLIGDLASLAFGSQRPLLDS 140
             :| || |   :  |      |:   |  |  ||   :           : : :  |:
Db      67 ATYILEAGKHRADKSTL---STYLEDCKFDRERIELFCTE----------YQNNKNSLEI 113

Qy     141 VAQQQGSSLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQLKL--TDGSAH---RFEVPIA 195
           :    | ||||::   ||::  | |:   |  :|: |: | :  ||  ::    |   :
Db     114 LLGSIGRSLPHITDVSWRLEYQIKTNQLHRMYRPAYLVTLSVQNTDSPSYPEISFSCSME 173

Qy     196 KFQELRYSVALVLKEMAELEKKCERKLQ 223
           : |:|      : :: :  |  ||  |
Db     174 QLQDL-------VGKLKDASKSLERATQ 194
```

RESULT 6
W24790
ID W24790 standard; Protein; 1786 AA.
AC W24790;
DT 08-OCT-1997 (first entry)
DE P. falciparum liver stage antigen-3.
KW Plasmodium falciparum; pre-erythrocyte; liver stage antigen; serum;
KW prophylaxis; Thai strain; gene organisation; exon; intron; hydrophobic;
KW glycosyl-phosphatidylinositol membrane anchoring sequence; antibody;

```
KW   vaccine; immunotherapy; malaria.
OS   Plasmodium falciparum.
FH   Key             Location/Qualifiers
FT   region          223. .278
FT                   /note= "repeat region 1"
FT   region          279. .818
FT                   /note= "repeat region 2"
FT   region          1537. .1576
FT                   /note= "repeat region 3"
PN   WO9641877-A2.
PD   27-DEC-1996.
PF   12-JUN-1996; F00894.
PR   13-JUN-1995; FR-007007.
PA   (INSP ) INST PASTEUR.
PI   Daubersies P, Druilhe P;
DR   WPI; 97-065464/06.
DR   N-PSDB; T78868.
PT   Plasmodium falciparum poly:peptide(s) and related nucleic acids -
PT   derived from the liver stage antigen-3, useful for malaria vaccine
PT   prodn. and diagnosis
PS   Claim 1; Fig 2A-I; 69pp; French.
CC   This sequence corresponds to a Plasmodium falciparum strain K1
CC   pre-erythrocytic liver stage antigen-3 (LSA-3) protein. The encoding
CC   gene sequence was isolated by screening a P. falciparum strain T9/96
CC   library with serum from a missionary treated by prophylaxis (for strain
CC   T6/96 see FR9101286). Of 20 clones isolated, clone 729S was used to
CC   screen a library generated from Thai strain K1. One clone contained a
CC   6.85 kb insert including the genomic sequence T78867. The gene comprises
CC   a 1.8 kb region encoding 3 major blocks of tetrapeptide repeats
CC   (especially the amino acid sequence VEES, VEEN, VEEI, VAPS, VAPT, etc)
CC   and a 3' hydrophobic region corresponding to a glycosyl-phosphatidyl-
CC   inositol membrane anchoring sequence. The invention relates to new
CC   polypeptides of at least 10 amino acids derived from the LSA-3 protein
CC   with the exception of the peptides W24791-4. The LSA-3 peptides can be
CC   used to raise antibodies and as vaccines for immunotherapy of malaria.
SQ   Sequence   1786 AA;

  Query Match             7.6%;  Score 84.5;  DB 1;  Length 1786;
  Best Local Similarity   21.6%;  Pred. No. 14;
  Matches   52;  Conservative   45;  Mismatches   79;  Indels   65;  Gaps   10;

Qy      27 SQPSPEVTAVAQLLKDLDRSTFRKLLKLVVGALHGKDCREAVEQLGASANLSEERLAVLL 86
           |  | | |   : :|| ::      : |||  :   |  |:||::    |: ||
Db    1183 SSDSKEET---ESIKDKEKD-----VSLVVEEVQDNDMDESVEKVLELKNMEEE------ 1229

Qy      87 AGTHTLLQQALRLPPASLKPDAFQEELQELG-IPQDLIGDLASL-------------AFG 132
             |:: |: :   : |     || |||  :  ||| |:  |
Db    1229 -----LMKDAVEINDITSK---LIEETQELNEVEADLIKDMEKLKELEKALSEDSKEIID 1280

Qy     133 SQRPLLDSVAQQQ---GSSLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQLK-------- 182
           ::   |: | :::    ::|  |   :   :  |   :  : |:  :| ::|
Db    1281 AKDDTLEKVIEEEHDITTTLDEVVELKDVEEDKIEKVSDLKDLEEDILKEVKEIKELESE 1340

Qy     182 -----------LTDGSAHRFEVP---IAKF----QELRYSVALVLKEMAELEKKCERKLQ 223
                      ||    : |:      ||    :|::   | :|||:: || : |:||:
Db    1341 ILEDYKELKTIETDILEEKKEIEKDHFEKFEEEAEEIKDLEADILKEVSSLEVEEEKKLE 1400

Qy     224 D 224
           :
Db    1401 E 1401


RESULT   7
W53924
ID   W53924 standard; peptide; 289 AA.
AC   W53924;
DT   20-AUG-1998  (first entry)
DE   Pseudomonas pseudoalcaligenes lipase variant.
KW   Lipase; variant; improved wash performance; removal; lipid stain;
KW   reduced calcium dependence; one-cycle wash efficiency; detergent;
KW   cleaning composition.
OS   Synthetic.
OS   Pseudomonas pseudoalcaligenes.
FH   Key             Location/Qualifiers
FT   Misc_difference 216
FT                   /label= V216P
FT                   /note= "Val optionally replaced with Pro"
FT   Misc_difference 224
FT                   /label= S224T
FT                   /note= "Ser optionally replaced with Thr"
FT   Misc_difference 229
FT                   /label= D229G
FT                   /note= "Asp optionally replaced with Gly"
FT   Misc_difference 231
FT                   /label= P231A
FT                   /note= "Pro optionally replaced with Ala"
FT   Misc_difference 242
FT                   /label= H242R
FT                   /note= "His optionally replaced with Arg"
FT   Misc_difference 245
FT                   /label= K245M
FT                   /note= "Lys optionally replaced with Met"
FT   Misc_difference 250
FT                   /label= D250N
FT                   /note= "Asp optionally replaced with Asn"
FT   Misc_difference 268
FT                   /label= L268I
FT                   /note= "Leu optionally replaced with Ile"
FT   Misc_difference 272
FT                   /label= D272S
FT                   /note= "Asp optionally replaced with Ser"
FT   Misc_difference 275
FT                   /label= T275S
FT                   /note= "Thr optionally replaced with Ser"
FT   Misc_difference 286
FT                   /label= L286X
FT                   /note= "Leu optionally replaced with Ile or Asn"
PN   WO9808939-A1.
PD   05-MAR-1998.
PF   26-AUG-1997; DK0345.
PR   29-AUG-1996; US-029190.
PR   27-AUG-1996; DK-000902.
PA   (NOVO ) NOVO-NORDISK AS.
PI   Okkels JS, Svendsen A;
DR   WPI; 98-230259/20.
PT   Variants of lipase from Pseudomonas containing specific amino acid
PT   substitutions - deletions or additions, having improved wash
PT   performance in detergent formulations
PS   Claim 5a; Page -; 104pp; English.
CC   The present sequence represents a variant of a wild-type Pseudomonas
CC   pseudoalcaligenes lipase. The variant lipase has have better wash
CC   performance than the original lipase. It has improved removal of
CC   lipid stains, reduced calcium dependence, better compatibility with
CC   detergents or their components, increased hydrophobicity, altered
CC   substrate specificity and better one-cycle wash efficiency. The variant
CC   lipase is used in detergent and cleaning compositions.
CC   note: this sequence does not appear in the specification; it was created
CC   using information provided.
SQ   Sequence   289 AA;

  Query Match             7.4%;  Score 82.5;  DB 1;  Length 289;
  Best Local Similarity   26.3%;  Pred. No. 1.9;
  Matches   49;  Conservative   23;  Mismatches   75;  Indels   39;  Gaps    9;

Qy      15 ADSHSGRVSFLG-SQPSPEVTAVAQLLKDLDRSTFRKLLKLVVGALH-GKDCREAVEQLG 72
             | |  |:|: :| |   | |  || :  ||  |         ||| | | |  : : |:
Db      74 AISGKGKVNLVGHSHGGPTVRYVAAVRPDLVASV------TSVGAPHKGSDTADFIRQI- 127

Qy      73 ASANLSEERLAVLLAGTHTLLQQALRLPPASLKPDAFQEELQELGIPQDLIGDLASL--- 130
             :  |  :| :: |   |:          |               ||: :| | ||
Db     127 PPGSAGEAIVAGIVNGLGALINFLSGSSSTS--------------PQNALGALESLKSE 171

Qy     130 ---AFGSQRP--LLDSVAQQQGSSLPHVSYFRWRVDVAISTSAQSRSLQPSVLM--QLKL 182
               || :: |  :  |   :    :  |||: |       ||  :  | || |:     |
```

```
Db    172 GAAAFNAKYPQGIPTSACGEGAYKVNGVSYYSWS-----GTSPLTNVLDPSDLLLGATSL 226

Qy    183 TDGSAH 188
          | | |:
Db    227 TFGEAN 232


RESULT   8
W46270
ID   W46270 standard; Protein; 758 AA.
AC   W46270;
DT   17-AUG-1998  (first entry)
DE   Moraxella catarrhalis strain 012E CopB outer membrane protein.
KW   CopB gene; outer membrane protein; epitope; infection;
KW   diagnosis; therapy; vaccine.
OS   Moraxella catarrhalis strain 012E.
FH   Key             Location/Qualifiers
FT   Region          10. .270
FT                   /note= "approximate location of conserved region"
FT   Region          275. .302
FT                   /note= "approximate location of non-conserved
FT                          region 1"
FT   Region          322. .329
FT                   /note= "approximate location of non-conserved
FT                          region 2"
FT   Region          348
FT                   /note= "approximate location of non-conserved
FT                          region 3"
FT   Region          350. .395
FT                   /note= "approximate location of conserved region"
FT   Region          400. .416
FT                   /note= "approximate location of non-conserved
FT                          region 4"
FT   Region          437. .450
FT                   /note= "approximate location of non-conserved
FT                          region 5"
FT   Region          455. .480
FT                   /note= "approximate location of conserved region"
FT   Region          485. .486
FT                   /note= "approximate location of non-conserved
FT                          region 6"
FT   Region          490. .520
FT                   /note= "approximate location of conserved region"
FT   Region          526. .541
FT                   /note= "approximate location of non-conserved
FT                          region 7"
FT   Region          545. .580
FT                   /note= "approximate location of conserved region"
FT   Region          583. .596
FT                   /note= "approximate location of non-conserved
FT                          region 8"
FT   Region          600. .635
FT                   /note= "approximate location of conserved region"
FT   Region          637. .640
FT                   /note= "approximate location of non-conserved
FT                          region 9"
FT   Region          645. .680
FT                   /note= "approximate location of conserved region"
FT   Region          683. .695
FT                   /note= "approximate location of non-conserved
FT                          region 10"
PN   WO9806851-A2.
PD   19-FEB-1998.
PF   12-AUG-1997; U14221.
PR   12-AUG-1996; US-023832.
PA   (TEXA ) UNIV TEXAS SYSTEM.
PI   Aebi C, Cope LD, Hansen EJ;
DR   WPI; 98-159542/14.
DR   N-PSDB; V26360.
PT   New isolated Moraxella catarrhalis peptide(s) - which define
PT   epitopes of the outer membrane protein used to develop products for
PT   the diagnosis, prophylaxis and treatment of infection
PS   Example 1; Page 91-94; 132pp; English.
CC   This protein comprises the CopB outer membrane protein of Moraxella
CC   catarrhalis strain 012E, encoded by an isolated copB gene (see
CC   V26360).  CopB represents an important antigenic determinant of M.
CC   catarrhalis, and a specific epitope (see W46250-51) has been
CC   identified in the non-conserved region 1 of strains 012E and
CC   035E that is bound by monoclonal antibody 10F3.  Claimed peptides
CC   (see W46274-81) of this region can be used in the diagnosis,
CC   treatment and prophylaxis (as vaccines) of M. catarrhalis
CC   infection.
SQ   Sequence   758 AA;


  Query Match             7.4%;  Score 82.5;  DB 1;  Length 758;
  Best Local Similarity   22.5%;  Pred. No. 6.9;
  Matches   39;  Conservative   28;  Mismatches   43;  Indels   63;  Gaps    8;

Qy     16 DSHSGRVSFLGSQPSPE------VTAVAQLLKDLDRS------TFRKLLKL--------- 55
          : ::|:   |||: :|:      | |  :|: ||||:      |::|| |
Db    295 NKYAGKGYKLGSK-TPDGKKYNVVDANGKLVADLDRNNPTQRETYQKLTNLEWTGKNLGF 353

Qy     55 -------VVGALHGKDC--------------REAVEQL-----------GASANLSEERL 82
                |   ||::                 :| :: :           ||: |  :|
Db    354 ANEVTANVYKLEHGRNSSSDQGNTYITKDVPKEIIDNVDTPSNMHVVATGANINFDKEFN 413

Qy     83 AVLLAG---THTLLQQALR------LPPASLKPDAFQEELQELGIPQDLIGDL 126
             | |    ||||:  :       :|| |||    :|  : ||  : :  :
Db    414 HSPLKGFGVDHTLLKYGINYRHQKAVPPRSLKPGVVHQEKTDAGIYLEAVNQI 466


RESULT   9
R75189
ID   R75189 standard; Protein; 1196 AA.
AC   R75189;
DT   30-MAY-1996  (first entry)
DE   Osteoinductive retrovirus RFB-14 pol gene product.
KW   RFB retrovirus; gag; pol; env; osteogenesis; osteoinductive protein;
KW   bone development; osteoporosis; gene therapy; polymerase;
KW   reverse transcriptase.
OS   Retrovirus RFB-14.
PN   DE4411718-A1.
PD   12-OCT-1995.
PF   05-APR-1994; 411718.
PR   05-APR-1994; DE-411718.
PA   (GSFU-) GSF FORSCHUNGSZENTRUM UMWELT & GESUNDHEI.
PI   Erfle V,  Gimbel W,  Oestergaard M,  Pedersen FS,  Pedersen L;
PI   Schmidt J, Strauss P;
DR   WPI; 95-352078/46.
DR   N-PSDB; Q94266.
PT   RFB-14 retrovirus genome  - and prodn. of osteo-inductive proteins
PS   Claim 14; Fig 1; 46pp; German.
CC   The full-length proviral genomic sequence of retrovirus RFB-14 has
CC   been determined. The virus codes for an osteoinductive protein,
CC   although the precise location of the coding region has not yet been
CC   identified. The virus may be useful in gene therapy of bone growth
CC   disorders such as osteoporosis. The present sequence is that of the
CC   viral pol gene product.
SQ   Sequence   1196 AA;


  Query Match             7.4%;  Score 82.5;  DB 1;  Length 1196;
  Best Local Similarity   24.2%;  Pred. No. 13;
  Matches   54;  Conservative   31;  Mismatches   83;  Indels   55;  Gaps   10;

Qy     33 VTAVAQLLKDLDRSTFRKLLKLVVGALHGKDCREAVEQLGASANLSEERL----AVLLAG 88
          | |:| | ||  : |  :   ||: | | :   ||: :      ||  |:    |:||
Db    533 VAAIAVLTKDAGKLTMGQ--PLVILAPHAE---EALVKQPPDRWLSNARMTHYQAMLLDT 587

Qy     89 THTLLQQALRLPPASLKPDAFQEELQELGIPQDLIGDLASLAFGSQRPLLDSV------- 142
                : | ||:| |      | | | | | :  ||    |::  | |
Db    588 DRVQFGPVVALNPATLLP------LPEEGAPHDCLEILAE-THGTRPDLTDQPIPDADHT 640

Qy    142 -----------AQQQGSSLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQLKLTDGSAHR 189
                    :: |:::   :   |   :   ||||   |  ::   ||: :|   |
```

```
Db    641 WYSDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSAQRAEL-IALTQALKMAEGK--R 697

Qy    190 FEVPIAKFQELRYSVA------------LVLKEMAELEKKCE 219
Db    698 LNV----YTDSRYAFATAHIHGEIYRRRGLLTSEGREIKNKSE 736


RESULT  10
P90476
ID   P90476 standard; peptide; 427 AA.
AC   P90476;
DT   24-OCT-1989  (first entry)
DE   Polypeptide of human Mullerian inhibiting substance.
KW   Mullerian inhibiting substance; human; cancer;
KW   contraception; N-terminal dimers.
OS   Homo sapiens (human).
FH   Key             Location/Qualifiers
PN   WO8906695-A.
PD   27-JUL-1989.
PF   25-JAN-1989; U00239.
PR   25-JAN-1988; US-281301.
PA   (BIOJ) Biogen. Inc.
PI   Cate R L; Pepinsky R B.
DR   WPI; 89-233849/32.
DR   N-PSDB; N90395.
PT   Cleaved dimers of Mullerian inhibiting substance like polypeptide(s)
PT   - are useful in treating cancers and for contraception.
PS   Claim 7(b); page 38; 64pp; English.
CC   Monomer of an N-terminal dimer of Mullerian inhibiting substance
CC   They are made by cleaving between Arg-Ser in the sequence
CC   aqrsag.  The dimers are used to treat cancer,
CC   esp. female genital tract cancer, and for contraception.  See also:
CC   N90395 and P90475-8.
SQ   Sequence   427 AA;


  Query Match             7.4%;  Score 82;  DB 1;  Length 427;
  Best Local Similarity   23.5%;  Pred. No. 3.6;
  Matches   54;  Conservative   17;  Mismatches   67;  Indels   92;  Gaps    9;

Qy     29 PSPEVTAVAQLL---KDLDRSTFRKLLKLVV----GALHGKDCREAVEQLGASANLSEER 81
Db    146 PGPEVTVTRAGLPGAQSLCPSRDTRYLVLAVDRPAGAWRGSGLALTLQPRGEDSRLSTAR 205

Qy     82 LAVLLAG---------------------------------------------------- 89
Db    206 LQALLFGDDHRCFTRMTPALLLLPRSEPAPLPAHGQLDTVPFPPPRPSAELEESPPSADP 265

Qy     89 ---THTLLQQALRLPPA-------SLKPDAFQEELQELGIPQDLI--GDLASL--AFGSQ 134
Db    266 FLETLTRLVRALRVPPARASAPRLALDPDAL------AGFPQGLVNLSDPAALERLLDGE 319

Qy    135 RPLL-----------DSVAQQQGSSLPHVSYFRWRVDVAI-STSAQSRSL 172
Db    320 EPLLLLLRPTAATTGDPAPLHDPTSAPWATALARRVAAELQAAAAELRSL 369


RESULT  11
R76502
ID   R76502 standard; Protein; 536 AA.
AC   R76502;
DT   19-DEC-1995  (first entry)
DE   Human MIS mature protein.
KW   Mullerian inhibiting substance; MIS; ovary cancer; therapy.
OS   Homo sapiens.
PN   US5427780-A.
PD   27-JUN-1995.
PF   30-OCT-1985; 792880.
PR   30-OCT-1985; US-792880.
PR   28-OCT-1986; US-923879.
PR   25-APR-1991; US-693764.
PR   05-OCT-1992; US-957061.
PA   (BIOJ ) BIOGEN INC.
PA   (GEHO ) GEN HOSPITAL CORP.
PI   Cate RL,  Donahoe PK;
DR   WPI; 95-239996/31.
PT   Compsn. comprising human Mullerian Inhibiting substance - useful for
PT   treating cancer, esp. ovarian cancer.
PS   Claim 1; Columns 7-8; 41pp; English.
CC   The insert of bovine MIS cDNA clone p521 was used to isolate the
CC   human MIS gene from a cosmid library.  The sequence of the encoded
CC   mature human MIS protein is given in R76502.
SQ   Sequence   536 AA;


  Query Match             7.4%;  Score 82;  DB 1;  Length 536;
  Best Local Similarity   23.5%;  Pred. No. 4.8;
  Matches   54;  Conservative   17;  Mismatches   67;  Indels   92;  Gaps    9;

Qy     29 PSPEVTAVAQLL---KDLDRSTFRKLLKLVV----GALHGKDCREAVEQLGASANLSEER 81
Db    146 PGPEVTVTRAGLPGAQSLCPSRDTRYLVLAVDRPAGAWRGSGLALTLQPRGEDSRLSTAR 205

Qy     82 LAVLLAG---------------------------------------------------- 89
Db    206 LQALLFGDDHRCFTRMTPALLLLPRSEPAPLPAHGQLDTVPFPPPRPSAELEESPPSADP 265

Qy     89 ---THTLLQQALRLPPA-------SLKPDAFQEELQELGIPQDLI--GDLASL--AFGSQ 134
Db    266 FLETLTRLVRALRVPPARASAPRLALDPDAL------AGFPQGLVNLSDPAALERLLDGE 319

Qy    135 RPLL-----------DSVAQQQGSSLPHVSYFRWRVDVAI-STSAQSRSL 172
Db    320 EPLLLLLRPTAATTGDPAPLHDPTSAPWATALARRVAAELQAAAAELRSL 369


RESULT  12
P70196
ID   P70196 standard; protein; 559 AA.
AC   P70196;
DT   10-APR-1991  (first entry)
DE   Sequence encoded by human mullerian inhibiting substance (MIS) gene
DE   in cosmid clone chmis33.
KW   Anti-tumour; cancer therapy; ovarian cancer; contraceptive.
OS   Homo sapiens.
PN   EP-221761-A.
PD   13-MAY-1987.
PF   29-OCT-1986; 308400.
PR   30-OCT-1985; US-792880.
PA   (BIOJ ) BIOGEN NV.
PA   (GEHO-) GEN HOSPITAL CORP.
PI   Cate RL, Donahoe PK;
DR   WPI; 87-158390/23.
DR   N-PSDB; N70307.
PT   DNA sequence encoding Mullerian inhibiting substance-like
PT   polypeptide - for use in treatment of cancer, esp. cancer of the
PT   female genital tract
PS   Example; Fig 6; 39pp; English.
CC   Bovine MIS protein from bovine testes was purified and the AA
CC   sequence of several fragments detd.  Based on these protein
CC   sequences, several antisense oligonucleotide DNA probes were
CC   synthesised.  The probes were then used to screen a bovine cDNA
CC   library.  The insert of one bovine MJS cDNA clone (pS21) was used
CC   to isolate the human MIS gene from a human cosmid library and
CC   partial cDNA clone from a human cDNA library.
SQ   Sequence   559 AA;


  Query Match             7.4%;  Score 82;  DB 1;  Length 559;
  Best Local Similarity   23.5%;  Pred. No. 5.1;
  Matches   54;  Conservative   17;  Mismatches   67;  Indels   92;  Gaps    9;

Qy     29 PSPEVTAVAQLL---KDLDRSTFRKLLKLVV----GALHGKDCREAVEQLGASANLSEER 81
Db    169 PGPEVTVTRAGLPGAQSLCPSRDTRYLVLAVDRPAGAWRGSGLALTLQPRGEDSRLSTAR 228
```

```
Qy      82 LAVLLAG---------------------------------------------------- 89
           |  || |
Db     229 LQALLFGDDHRCFTRMTPALLLLPRSEPAPLPAHGQLDTVPFPPPRPSAELEESPPSADP 288

Qy      89 ---THTLLQQALRLPPA-------SLKPDAFQEELQELGIPQDLI--GDLASL--AFGSQ 134
             | | | :|||:|||       :| |||        | || |:   | |:|      :
Db     289 FLETLTRLVRALRVPPARASAPRLALDPDAL------AGFPQGLVNLSDPAALERLLDGE 342

Qy     135 RPLL-----------DSVAQQQGSSLPHVSYFRWRVDVAI-STSAQSRSL 172
            |||           |       :| |  :    ||   : : :|: |||
Db     343 EPLLLLLRPTAATTGDPAPLHDPTSAPWATALARRVAAELQAAAAELRSL 392


RESULT  13
R76501
ID   R76501 standard; Protein; 560 AA.
AC   R76501;
DT   19-DEC-1995  (first entry)
DE   Human MIS protein.
KW   Mullerian inhibiting substance; MIS; ovary cancer; therapy.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   protein         25. .536
FT                   /label= Mat_protein
FT                   /note= "Claim 1, column 26"
PN   US5427780-A.
PD   27-JUN-1995.
PF   30-OCT-1985; 792880.
PR   30-OCT-1985; US-792880.
PR   28-OCT-1986; US-923879.
PR   25-APR-1991; US-693764.
PR   05-OCT-1992; US-957061.
PA   (BIOJ ) BIOGEN INC.
PA   (GEHO ) GEN HOSPITAL CORP.
PI   Cate RL,  Donahoe PK;
DR   WPI; 95-239996/31.
DR   N-PSDB; Q92785.
PT   Compsn. comprising human Mullerian Inhibiting substance - useful for
PT   treating cancer, esp. ovarian cancer.
PS   Disclosure; Fig.6a-6k; 41pp; English.
CC   The insert of bovine MIS cDNA clone p521 was used to isolate the
CC   human MIS gene from a cosmid library.  The DNA can be operatively
CC   linked to expression control sequences and used in mammalian or
CC   other host cells to produce MIS-like polypeptides.
SQ   Sequence   560 AA;


  Query Match             7.4%;  Score 82;  DB 1;  Length 560;
  Best Local Similarity   23.5%;  Pred. No. 5.1;
  Matches   54;  Conservative   17;  Mismatches   67;  Indels   92;  Gaps   9;

Qy      29 PSPEVTAVAQLL---KDLDRSTFRKLLKLVV----GALHGKDCREAVEQLGASANLSEER 81
           | ||||     |   : |  |   : | | |    ||  |      ::  |  : ||  |
Db     170 PGPEVTVTRAGLPGAQSLCPSRDTRYLVLAVDRPAGAWRGSGLALTLQPRGEDSRLSTAR 229

Qy      82 LAVLLAG---------------------------------------------------- 89
           |  || |
Db     230 LQALLFGDDHRCFTRMTPALLLLPRSEPAPLPAHGQLDTVPFPPPRPSAELEESPPSADP 289

Qy      89 ---THTLLQQALRLPPA-------SLKPDAFQEELQELGIPQDLI--GDLASL--AFGSQ 134
             | | | :|||:|||       :| |||        | || |:   | |:|      :
Db     290 FLETLTRLVRALRVPPARASAPRLALDPDAL------AGFPQGLVNLSDPAALERLLDGE 343

Qy     135 RPLL-----------DSVAQQQGSSLPHVSYFRWRVDVAI-STSAQSRSL 172
            |||           |       :| |  :    ||   : : :|: |||
Db     344 EPLLLLLRPTAATTGDPAPLHDPTSAPWATALARRVAAELQAAAAELRSL 393


RESULT  14
R34712
ID   R34712 standard; Protein; 3588 AA.
AC   R34712;
DT   17-AUG-1993  (first entry)
DE   Bacillus subtilis srfA operon ORF1 prod.
KW   Mulienzyme complex; surfactin synthetase; MCSS; ORF; surfactant.
OS   Bacillus subtilis.
FH   Key             Location/Qualifiers
FT   peptide         139. .145
FT                   /note= "repeat sequence"
FT   peptide         1184. .1190
FT                   /note= "repeat sequence"
FT   peptide         2223. .2229
FT                   /note= "repeat sequence"
FT   peptide         3255. .3261
FT                   /note= "repeat sequence"
FT   peptide         606. .616
FT                   /note= "repeat sequence"
FT   peptide         1653. .1663
FT                   /note= "repeat sequence"
FT   peptide         2698. .2709
FT                   /note= "repeat sequence"
FT   peptide         858. .880
FT                   /note= "repeat sequence"
FT   peptide         1897. .1919
FT                   /note= "repeat sequence"
FT   peptide         2930. .2949
FT                   /note= "repeat sequence"
PN   EP-540074-A.
PD   05-MAY-1993.
PF   03-OCT-1992; 203037.
PR   09-OCT-1991; IT-MI2683.
PR   02-SEP-1992; IT-MI2044.
PA   (ENIE ) ENIRICERCHE SPA.
PI   Carrera P,  Cosmina P,  De FERRA F,  Grandi G,  Perego M;
PI   Rodriguez F;
DR   WPI; 93-145447/18.
DR   N-PSDB; Q40706.
PT   Multi-enzyme complex surfactin synthetase DNA - is isolated from
PT   Bacillus subtilis, and used for prodn. of surfactin for use in
PT   pharmaceutical prods.
PS   Claim 15; Page 27-34; 70pp; English.
CC   The Bacillus subtilis chromosomal DNA region comprises the srfA
CC   operon which encodes the multienzyme complex surfactin synthetase
CC   (MCSS).. Analysis of the sequence showed four regions potentially
CC   coding for proteins, a zone upstream of the first ORF contg. the
CC   srfA operon promoter and a presumed terminator positioned downstream
CC   of the stop codon of the fourth ORF.  ORF1 encodes a protein
CC   (shown) which can be divided into 3 adjacent repeat regions with an
CC   internal homology of ca. 1000 bases, more marked in the second half of
CC   these regions.  At the C terminal end of module 3 there is a region
CC   (module A) of 500 amino acids which has no homology with modules 1-3.
CC   Although ORF1 codes for a protein of unknown function, regions of the
CC   protein are found to be highly homlogous to synthetases coded by TycA
CC   and GrsA (tyrocidin and gramicidin synthetase subunits I) and with
CC   aminoadipyl cystein valine synthetase.
CC   See also R34713-21.
SQ   Sequence   3588 AA;


  Query Match             7.3%;  Score 81;  DB 1;  Length 3588;
  Best Local Similarity   20.7%;  Pred. No. 80;
  Matches   56;  Conservative   36;  Mismatches   87;  Indels   92;  Gaps   11;

Qy       1 MSALGAAAAPYLHHPAD--------SHSGRVSFLGSQPSPEVTAVAQLLKDLDRSTFRKLL 52
           :  |    | |  | |           |::|  |  : :  |: :| |: | : :  ||
Db    2279 LKELDGELPVLTLPTDYSRPAVQTFEGDRIAF--SLEAAKADALRRLAKETDSTLYMVLL 2336

Qy      53 -------------KLVVGA-LHGKDCREAVEQLGASANLSEERLAVLLAGTHTLLQQAL 97
                        :  ::||: : |:   :    :|   |              ||   ||
Db    2337 ASYSAFLSKICGQDDIIVGSPVAGRSQADVSRVIGMFVN-------------TL---AL 2379

Qy      98 RLPPASLKPDA-FQEELQELGIP---------QDLIGDLASLAFGSQRPLLDSVAQQQGS 147
           |  |   |  | :  |::|  :          :||||::      |  || |:|   | :
Db    2380 RTYPKGEKTFADYLNEVKETALSAFHAQDYPLEDLIGNVQVQRDTSSNPLFDAVFSMQNA 2439

Qy     148 SLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQLKLTDGSAHRFEVPIAKFQ--------- 199
```

```
          ::  :             : :   |:|            | ||   |||
Db   2440 NIKDL------------TMKGIQLEP------------HPFERKTAKFDLTLTADETD 2473

Qy    199 -----ELRYSVALVLKEMAELEKKCERKLQD 224
                | |: ||  :|  |  |:   :| |
Db   2474 GGLTFVLEYNTALFKQETIERWKQYWMELLD 2504


RESULT  15
R88023
ID   R88023 standard; protein; 289 AA.
AC   R88023;
DT   02-AUG-1996  (first entry)
DE   Mature Pseudomonas pseudoalcaligenes lipase L210R variant.
KW   Mutant; lipase; enzymatic detergent; substitution; variant; improved;
KW   hydrophobicity.
OS   Pseudomonas pseudoalcaligenes.
FH   Key             Location/Qualifiers
FT   misc_difference 210
FT                   /label= substitution
FT                   /note= "Leu to Arg, in wild-type sequence a Leu
FT                   residue is present at position 210 of mature
FT                   P. pseudoalcaligenes lipase"
FT   misc_difference 27
FT                   /note= "possible site for Ile to Arg substitution"
FT   misc_difference 129
FT                   /note= "possible site for Gly to Tyr substitution"
FT   misc_difference 135
FT                   /note= "possible site for Ile to Arg substitution"
FT   misc_difference 155
FT                   /note= "possible site for Ser to Arg substitution"
FT   misc_difference 213
FT                   /note= "possible site for Val to Phe substitution"
FT   misc_difference 262
FT                   /note= "possible site for Thr to Phe substitution"
PN   WO9535381-A1.
PD   28-DEC-1995.
PF   15-JUN-1995; E02349.
PR   20-JUN-1994; EP-201761.
PA   (UNIL ) UNILEVER NV.
PA   (UNIL ) UNILEVER PLC.
PI   De Vlieg J,  Frenken LGJ,  Peters H,  Suerbaum HMU;
PI   Verrips CT;
DR   WPI; 96-058418/06.
PT   Pseudomonas lipase variant with increased surface hydrophobicity  -
PT   has improved in-the-wash performance, useful in enzymatic detergent
PS   Disclosure; Page -; 33pp; English.
CC   R88019-R88025 are protein variants of Pseudomonas pseudoalcaligenes
CC   lipase. The sequences are based upon the wild-type P. pseudoalcaligenes
CC   sequence disclosed in WO9402617-A (Gist-Brocades). The lipase variants
CC   may contain one or more of the amino acid substitutions indicated in
CC   the features table. Using these variants it was shown that lipases
CC   can be modified in a such a way that interaction with the substrate
CC   can be improved without forming large hydrophobic areas on the modified
CC   lipase surface which allow aggregation of lipase molecules. The
CC   lipase variants show improved in-the-wash lipolytic activity and may
CC   be used in enzymatic detergent compsns.
SQ   Sequence   289 AA;


  Query Match             7.2%;  Score 80.5;  DB 1;  Length 289;
  Best Local Similarity   25.9%;  Pred. No. 2.9;
  Matches   49;  Conservative   24;  Mismatches   75;  Indels   41;  Gaps    9;

Qy      15 ADSHSGRVSFLG-SQPSPEVTAVAQLLKDLDRSTFRKLLKLVVGALH-GKDCREAVEQLG 72
           | |  |:|: :| |   | |  || :  ||  |          ||| | | |  : : |:
Db      74 AISGKGKVNLVGHSHGGPTVRYVAAVRPDLVASV------TSVGAPHKGSDTADFIRQI- 127

Qy      73 ASANLSEERLAVLLAGTHTLLQQALRLPPASLKPDAFQEELQELGIPQDLIGDLASL--- 130
             :  | |:| :: |   |:          |               ||: :| | ||
Db     127 PPYSAGEARVAGIVNGLGALINFLSGSSSTS--------------PQNALGALESLNSE 171

Qy     130 ---AFGSQRP--LLDSVAQQQGSSLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQLKLTD 184
                 || :: |  :  |   :    : |||: |      || ::  |  | |   |
Db     172 GAAAFNAKYPQGIPTSACGEGAYKVNGVSYYSWS-----GTSPRTNVLDVSDL---LLG 222

Qy     185 GSAHRFEVP 193
             |:  |: |
Db     223 ASSLTFDEP 231


Search completed: May 15, 2000, 02:06:06
Job time: 10076 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: May 15, 2000, 01:46:14 ; Search time 35.18 Seconds
(without alignments)
92.094 Million cell updates/sec

Title: US-09-223-796-2
Perfect score: 1112
Sequence: 1 MSALGAAAPYLHHPADSHSG.....ALVLKEMAELEKKCERKLQD 224

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 143561 seqs, 14463640 residues

Total number of hits satisfying chosen parameters: 143561

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	86.5	7.8	195	2	US-08-822-260-1	Sequence 1, Appli
2	82.5	7.4	195	2	US-08-822-260-3	Sequence 3, Appli
3	82	7.4	560	1	US-07-683-957B-1	Sequence 1, Appli
4	79.5	7.1	681	1	US-08-083-590A-18	Sequence 18, Appl
5	79.5	7.1	681	2	US-08-346-128-37	Sequence 37, Appl
6	79.5	7.1	1078	1	US-08-264-534-32	Sequence 32, Appl
7	79.5	7.1	1078	1	US-08-083-590A-11	Sequence 11, Appl
8	79.5	7.1	1078	1	US-08-465-500-32	Sequence 32, Appl
9	79.5	7.1	1078	2	US-08-346-128-32	Sequence 32, Appl
10	79.5	7.1	2556	1	US-08-083-590A-20	Sequence 20, Appl
11	78.5	7.1	363	1	US-08-530-950-6	Sequence 6, Appli
12	78.5	7.1	399	1	US-08-530-950-10	Sequence 10, Appl
13	78.5	7.1	399	2	US-08-874-186-92	Sequence 92, Appl
14	77	6.9	3567	2	US-07-642-734C-4	Sequence 4, Appli
15	76.5	6.9	393	1	US-08-530-950-8	Sequence 8, Appli
16	76	6.8	689	1	US-08-221-817-18	Sequence 18, Appl
17	76	6.8	689	1	US-08-454-439-18	Sequence 18, Appl
18	76	6.8	689	4	PCT-US94-10487-18	Sequence 18, Appl
19	75.5	6.8	226	3	US-08-966-318-1	Sequence 1, Appli
20	75.5	6.8	226	3	US-08-966-318-1	Sequence 1, Appli
21	75	6.7	635	2	US-08-484-101B-36	Sequence 36, Appl
22	75	6.7	635	2	US-08-484-101B-50	Sequence 50, Appl
23	74.5	6.7	334	5	5290690-10	Patent No. 5290690
24	74.5	6.7	335	5	5290690-9	Patent No. 5290690
25	74.5	6.7	1079	2	US-08-929-967-8	Sequence 8, Appli
26	74	6.7	688	1	US-08-221-817-19	Sequence 19, Appl
27	74	6.7	688	1	US-08-454-439-19	Sequence 19, Appl
28	74	6.7	688	4	PCT-US94-10487-19	Sequence 19, Appl
29	74	6.7	1111	1	US-08-317-450B-15	Sequence 15, Appl
30	74	6.7	1163	1	US-08-173-497-4	Sequence 4, Appli
31	74	6.7	1163	1	US-08-286-889-4	Sequence 4, Appli
32	74	6.7	1163	1	US-08-485-618-4	Sequence 4, Appli
33	74	6.7	1163	1	US-08-362-652-4	Sequence 4, Appli
34	74	6.7	1163	2	US-08-605-672-4	Sequence 4, Appli
35	74	6.7	1163	2	US-08-482-293A-4	Sequence 4, Appli
36	74	6.7	1163	2	US-08-943-363-4	Sequence 4, Appli
37	74	6.7	1163	2	US-08-476-062A-44	Sequence 44, Appl
38	74	6.7	1163	4	PCT-US96-01314-44	Sequence 44, Appl
39	74	6.7	1193	1	US-08-317-450B-13	Sequence 13, Appl
40	74	6.7	3491	2	US-07-642-734C-2	Sequence 2, Appli
41	73.5	6.6	668	1	US-08-205-018-2	Sequence 2, Appli
42	73.5	6.6	859	1	US-08-395-580-2	Sequence 2, Appli
43	73.5	6.6	859	4	PCT-US95-02792-2	Sequence 2, Appli
44	72.5	6.5	532	1	US-08-188-228-44	Sequence 44, Appl
45	72.5	6.5	532	1	US-08-332-638-44	Sequence 44, Appl

ALIGNMENTS

```
RESULT 1
US-08-822-260-1
; Sequence 1, Application US/08822260
; Patent No. 5830660
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Goli, Surya K.
;    TITLE OF INVENTION:  NOVEL TUMORIGENESIS PROTEIN
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/822,260
;      FILING DATE:  Herewith
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0247 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  195 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  UTRSNOT02
;      CLONE:  2267574
US-08-822-260-1


  Query Match             7.8%;  Score 86.5;  DB 2;  Length 195;
  Best Local Similarity   22.1%;  Pred. No. 0.067;
  Matches   46;  Conservative   33;  Mismatches   82;  Indels   47;  Gaps    8;
```

```
Qy      38 QLLKD---LDRSTFRKLLKLVVGALHGKDCREAV----------EQLGASANLSEE 80
           |:| |    | : |  ||:    :|      |||             :   |:|
Db      12 QMLADPRSFDSNAFTLLLRAAFQSLLDAQADEAVLDHPDLKHIDPVVLKHCHAAA----- 67

Qy      81 RLAVLLAGTHTLLQQALRLPPASLKPDAFQEELQELGIPQDLIGDLASLAFGSQRPLLDS 140
             :| || |   :  |      |:   |  |  ||   :          : : :  |:
Db      67 ATYILEAGKHRADKSTL---STYLEDCKFDRERIELFCTE----------YQNNKNSLEI 113

Qy     141 VAQQQGSSLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQLKL--TDGSAH---RFEVPIA 195
           :   | ||||::   ||::  | |:   |  :|: |: | :  ||  ::    |   :
Db     114 LLGSIGRSLPHITDVSWRLEYQIKTNQLHRMYRPAYLVTLSVQNTDSPSYPEISFSCSME 173

Qy     196 KFQELRYSVALVLKEMAELEKKCERKLQ 223
           : |:|       : :: :  |  ||  |
Db     174 QLQDL-------VGKLKDASKSLERATQ 194


RESULT   2
US-08-822-260-3
; Sequence 3, Application US/08822260
; Patent No. 5830660
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Goli, Surya K.
;    TITLE OF INVENTION:  NOVEL TUMORIGENESIS PROTEIN
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/822,260
;      FILING DATE:  Herewith
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0247 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  195 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  GenBank
;      CLONE:  265569
US-08-822-260-3


  Query Match             7.4%;  Score 82.5;  DB 2;  Length 195;
  Best Local Similarity   20.0%;  Pred. No. 0.18;
  Matches   29;  Conservative   31;  Mismatches   60;  Indels   25;  Gaps    4;

Qy      84 VLLAGTHTLLQQALRLPPASLKPDAFQEELQELGIPQDLIGDLASLAFGSQRPLLDSVAQ 143
           :| || | : :  |      |:   |  |  ||   :          : : :  |:::
Db      70 ILEAGKHQVDKSTL---STYLEDCKFDRERIELFCTE----------YQNNKNSLETLLG 116

Qy     144 QQGSSLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQLKLTDGSAHRFEVPIAKFQELRYS 203
           | ||||::   ||::  | |:   :  :|  |: | : :  :         : |: :|
Db     117 SIGRSLPHITDVSWRLEYQIKTNQLHKMYRPGYLVTLNVENNDSQ-------SYPEINFS 169

Qy     204 VAL-----VLKEMAELEKKCERKLQ 223
           :     :: :: :  |  ||  |
Db     170 CNMEQLQDLVGKLKDASKSLERATQ 194


RESULT   3
US-07-683-957B-1
; Sequence 1, Application US/07683957B
; Patent No. 5310880
;  GENERAL INFORMATION:
;    APPLICANT:  Donahoe, Patricia K.
;    APPLICANT:  Ragin, Richard C.
;    APPLICANT:  MacLaughlin, David T.
;    TITLE OF INVENTION:  Purification of M llerian Inhibiting
;    TITLE OF INVENTION:  Substance
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox
;      STREET:  1100 New York Avenue, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/683,957B
;      FILING DATE:  19910412
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Goldstein, Jorge A.
;      REGISTRATION NUMBER:  29,021
;      REFERENCE/DOCKET NUMBER:  0609.3060000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 466-0800
;      TELEFAX:  (202) 833-8716
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  560 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-683-957B-1


  Query Match             7.4%;  Score 82;  DB 1;  Length 560;
  Best Local Similarity   23.5%;  Pred. No. 1.1;
  Matches   54;  Conservative   17;  Mismatches   67;  Indels   92;  Gaps    9;

Qy      29 PSPEVTAVAQLL---KDLDRSTFRKLLKLVV----GALHGKDCREAVEQLGASANLSEER 81
           | ||||     |   : |  |   : | | |    ||  |      ::  |  : ||  |
Db     170 PGPEVTVTRAGLPGAQSLCPSRDTRYLVLAVDRPAGAWRGSGLALTLQPRGEDSRLSTAR 229

Qy      82 LAVLLAG---------------------------------------------------- 89
           |  || |
Db     230 LQALLFGDDHRCFTRMTPALLLLPRSEPAPLPAHGQLDTVPFPPPPRPSAELEESPPSADP 289

Qy      89 ---THTLLQQALRLPPA-------SLKPDAFQEELQELGIPQDLI--GDLASL--AFGSQ 134
             | | | :|||:|||       :| |||       | || |:   | |:|     :
Db     290 FLETLTRLVRALRVPPARASAPRLALDPDAL------AGFPQGLVNLSDPAALERLLDGE 343

Qy     135 RPLL-----------DSVAQQQGSSLPHVSYFRWRVDVAI-STSAQSRSL 172
           |||            |  :| | :   ||   : : :|: |||
Db     344 EPLLLLLRPTAATTGDPAPLHDPTSAPWATALARRVAAELQAAAAELRSL 393
```

```
RESULT   4
US-08-083-590A-18
; Sequence 18, Application US/08083590A
; Patent No. 5786158
;  GENERAL INFORMATION:
;    APPLICANT:  Artavanis-Tsakonas, S. et al.
;    TITLE OF INVENTION:  Therapeutic And Diagnostic Methods
;    TITLE OF INVENTION:  And Compositions Based On No. 5786158ch Proteins And
;    TITLE OF INVENTION:  Nucleic Acids
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/083,590A
;      FILING DATE:  25-JUN-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  7326-015
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 790-9090
;      TELEFAX:  212 8698864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  681 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-083-590A-18


  Query Match             7.1%;  Score 79.5;  DB 1;  Length 681;
  Best Local Similarity   28.9%;  Pred. No. 2.7;
  Matches   44;  Conservative   17;  Mismatches   76;  Indels   15;  Gaps    6;

Qy      55 VVGALHGKDCREAVEQLGASANLSEERLAVL--LAGTHTLLQQALRLPPASLKPDAFQEE 112
           :|| ||       |: |: :   |   |||    |  |  :  | |::   :|:|   |::
Db     486 MVGPLHSSLAASALSQMMSYQGLPSTRLATQPHLVQTQQVQPQNLQMQQQNLQPANIQQQ 545

Qy     113 --LQ--------ELGIPQDLIGDLASLAFGSQRPLLDSVAQQQGSSLPHVSYFRWRVDVA 162
             ||          ||:     | |   :| | |    | | | |    |      :   |
Db     546 QSLQPPPPPPQPHLGVSSAASGHLGR-SFLSGEPSQADV-QPLGPSSLAVHTILPQESPA 603

Qy     163 ISTSAQSRSLQPSVLMQLKLTDGSAHRFEVPI 194
           : ||  | || | |     ||  | | :  |:
Db     604 LPTSLPS-SLVPPVTAAQFLTPPSQHSYSSPV 634


RESULT   5
US-08-346-128-37
; Sequence 37, Application US/08346128
; Patent No. 5856441
;  GENERAL INFORMATION:
;    APPLICANT:  Artavanis-Tsakonas, Spyridon et al.
;    TITLE OF INVENTION:  Human No. 5856441ch And Delta, Binding Domains
;    TITLE OF INVENTION:  In Toporythmic Proteins, And Methods Based Thereon
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/346,128
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/879,038
;      FILING DATE:  30-APR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  7326-009
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 790-9090
;      TELEFAX:  212 8698864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  37:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  681 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-346-128-37


  Query Match             7.1%;  Score 79.5;  DB 2;  Length 681;
  Best Local Similarity   28.9%;  Pred. No. 2.7;
  Matches   44;  Conservative   17;  Mismatches   76;  Indels   15;  Gaps    6;

Qy      55 VVGALHGKDCREAVEQLGASANLSEERLAVL--LAGTHTLLQQALRLPPASLKPDAFQEE 112
           :|| ||       |: |: :   |   |||    |  |  :  | |::   :|:|   |::
Db     486 MVGPLHSSLAASALSQMMSYQGLPSTRLATQPHLVQTQQVQPQNLQMQQQNLQPANIQQQ 545

Qy     113 --LQ--------ELGIPQDLIGDLASLAFGSQRPLLDSVAQQQGSSLPHVSYFRWRVDVA 162
             ||          ||:     | |   :| | |    | | | |    |      :   |
Db     546 QSLQPPPPPPQPHLGVSSAASGHLGR-SFLSGEPSQADV-QPLGPSSLAVHTILPQESPA 603

Qy     163 ISTSAQSRSLQPSVLMQLKLTDGSAHRFEVPI 194
           : ||  | || | |     ||  | | :  |:
Db     604 LPTSLPS-SLVPPVTAAQFLTPPSQHSYSSPV 634


RESULT   6
US-08-264-534-32
; Sequence 32, Application US/08264534
; Patent No. 5648464
;  GENERAL INFORMATION:
;    APPLICANT:  Artavanis-Tsakonas, Spyridon et al.
;    TITLE OF INVENTION:  Human No. 5648464ch And Delta, Binding Domains
;    TITLE OF INVENTION:  In Toporythmic Proteins, And Methods Based Thereon
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/08/264,534
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/695,189
;      FILING DATE:  03-MAY-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  7326-004
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 790-9090
;      TELEFAX:  212 8698864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1078 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-264-534-32


  Query Match             7.1%;  Score 79.5;  DB 1;  Length 1078;
  Best Local Similarity   28.9%;  Pred. No. 5.5;
  Matches   44;  Conservative   17;  Mismatches   76;  Indels   15;  Gaps    6;

Qy      55 VVGALHGKDCREAVEQLGASANLSEERLAVL--LAGTHTLLQQALRLPPASLKPDAFQEE 112
           :|| ||      |: |: :   |   |||    |  |  :  | |::   :|:|   |::
Db     893 MVGPLHSSLAASALSQMMSYQGLPSTRLATQPHLVQTQQVQPQNLQMQQQNLQPANIQQQ 952

Qy     113 --LQ--------ELGIPQDLIGDLASLAFGSQRPLLDSVAQQQGSSLPHVSYFRWRVDVA 162
             ||          ||:     | |   :| |  |    | |  | |  |     :  |
Db     953 QSLQPPPPPPQPHLGVSSAASGHLGR-SFLSGEPSQADV-QPLGPSSLAVHTILPQESPA 1010

Qy     163 ISTSAQSRSLQPSVLMQLKLTDGSAHRFEVPI 194
           : ||  | || | |     ||  | | :  |:
Db    1011 LPTSLPS-SLVPPVTAAQFLTPPSQHSYSSPV 1041


RESULT   7
US-08-083-590A-11
; Sequence 11, Application US/08083590A
; Patent No. 5786158
;  GENERAL INFORMATION:
;    APPLICANT:  Artavanis-Tsakonas, S. et al.
;    TITLE OF INVENTION:  Therapeutic And Diagnostic Methods
;    TITLE OF INVENTION:  And Compositions Based On No. 5786158ch Proteins And
;    TITLE OF INVENTION:  Nucleic Acids
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/083,590A
;      FILING DATE:  25-JUN-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  7326-015
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 790-9090
;      TELEFAX:  212 8698864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1078 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-083-590A-11


  Query Match             7.1%;  Score 79.5;  DB 1;  Length 1078;
  Best Local Similarity   28.9%;  Pred. No. 5.5;
  Matches   44;  Conservative   17;  Mismatches   76;  Indels   15;  Gaps    6;

Qy      55 VVGALHGKDCREAVEQLGASANLSEERLAVL--LAGTHTLLQQALRLPPASLKPDAFQEE 112
           :|| ||      |: |: :   |   |||    |  |  :  | |::   :|:|   |::
Db     893 MVGPLHSSLAASALSQMMSYQGLPSTRLATQPHLVQTQQVQPQNLQMQQQNLQPANIQQQ 952

Qy     113 --LQ--------ELGIPQDLIGDLASLAFGSQRPLLDSVAQQQGSSLPHVSYFRWRVDVA 162
             ||          ||:     | |   :| |  |    | |  | |  |     :  |
Db     953 QSLQPPPPPPQPHLGVSSAASGHLGR-SFLSGEPSQADV-QPLGPSSLAVHTILPQESPA 1010

Qy     163 ISTSAQSRSLQPSVLMQLKLTDGSAHRFEVPI 194
           : ||  | || | |     ||  | | :  |:
Db    1011 LPTSLPS-SLVPPVTAAQFLTPPSQHSYSSPV 1041


RESULT   8
US-08-465-500-32
; Sequence 32, Application US/08465500
; Patent No. 5789195
;  GENERAL INFORMATION:
;    APPLICANT:  Artavanis-Tsakonas, Spyridon
;    APPLICANT:  Muskavitch, Marc A.T.
;    APPLICANT:  Fehon, Richard G.
;    APPLICANT:  Rebay, Ilaria
;    APPLICANT:  Blaumueller, Cristine M.
;    APPLICANT:  Shepard, Scott B.
;    TITLE OF INVENTION:  HUMAN NOTCH AND DELTA, BINDING DOMAINS
;    TITLE OF INVENTION:  IN TOPORYTHMIC PROTEINS,AND METHODS BASED THEREON
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/465,500
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  7326-034
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864/9741
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1078 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-465-500-32
```

```
  Query Match             7.1%;  Score 79.5;  DB 1;  Length 1078;
  Best Local Similarity   28.9%;  Pred. No. 5.5;
  Matches   44;  Conservative   17;  Mismatches   76;  Indels   15;  Gaps    6;

Qy      55 VVGALHGKDCREAVEQLGASANLSEERLAVL--LAGTHTLLQQALRLPPASLKPDAFQEE 112
           :|| ||      |: |: :   |   |||    |  |  :  | |::   :|:|   |::
Db     893 MVGPLHSSLAASALSQMMSYQGLPSTRLATQPHLVQTQQVQPQNLQMQQQNLQPANIQQQ 952

Qy     113 --LQ--------ELGIPQDLIGDLASLAFGSQRPLLDSVAQQQGSSLPHVSYFRWRVDVA 162
             ||         ||:     | |   :| |  |    | |  | |   |     :   |
Db     953 QSLQPPPPPPQPHLGVSSAASGHLGR-SFLSGEPSQADV-QPLGPSSLAVHTILPQESPA 1010

Qy     163 ISTSAQSRSLQPSVLMQLKLTDGSAHRFEVPI 194
           : ||  | || | |     ||  | | :  |:
Db    1011 LPTSLPS-SLVPPVTAAQFLTPPSQHSYSSPV 1041


RESULT  9
US-08-346-128-32
; Sequence 32, Application US/08346128
; Patent No. 5856441
;  GENERAL INFORMATION:
;    APPLICANT:  Artavanis-Tsakonas, Spyridon et al.
;    TITLE OF INVENTION:  Human No. 5856441ch And Delta, Binding Domains
;    TITLE OF INVENTION:  In Toporythmic Proteins, And Methods Based Thereon
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/346,128
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/879,038
;      FILING DATE:  30-APR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  7326-009
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 790-9090
;      TELEFAX:  212 8698864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1078 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-346-128-32


  Query Match             7.1%;  Score 79.5;  DB 2;  Length 1078;
  Best Local Similarity   28.9%;  Pred. No. 5.5;
  Matches   44;  Conservative   17;  Mismatches   76;  Indels   15;  Gaps    6;

Qy      55 VVGALHGKDCREAVEQLGASANLSEERLAVL--LAGTHTLLQQALRLPPASLKPDAFQEE 112
           :|| ||      |: |: :   |   |||    |  |  :  | |::   :|:|   |::
Db     893 MVGPLHSSLAASALSQMMSYQGLPSTRLATQPHLVQTQQVQPQNLQMQQQNLQPANIQQQ 952

Qy     113 --LQ--------ELGIPQDLIGDLASLAFGSQRPLLDSVAQQQGSSLPHVSYFRWRVDVA 162
             ||         ||:     | |   :| |  |    | |  | |   |     :   |
Db     953 QSLQPPPPPPQPHLGVSSAASGHLGR-SFLSGEPSQADV-QPLGPSSLAVHTILPQESPA 1010

Qy     163 ISTSAQSRSLQPSVLMQLKLTDGSAHRFEVPI 194
           : ||  | || | |     ||  | | :  |:
Db    1011 LPTSLPS-SLVPPVTAAQFLTPPSQHSYSSPV 1041


RESULT  10
US-08-083-590A-20
; Sequence 20, Application US/08083590A
; Patent No. 5786158
;  GENERAL INFORMATION:
;    APPLICANT:  Artavanis-Tsakonas, S. et al.
;    TITLE OF INVENTION:  Therapeutic And Diagnostic Methods
;    TITLE OF INVENTION:  And Compositions Based On No. 5786158ch Proteins And
;    TITLE OF INVENTION:  Nucleic Acids
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/083,590A
;      FILING DATE:  25-JUN-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  7326-015
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 790-9090
;      TELEFAX:  212 8698864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2556 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-083-590A-20


  Query Match             7.1%;  Score 79.5;  DB 1;  Length 2556;
  Best Local Similarity   28.9%;  Pred. No. 21;
  Matches   44;  Conservative   17;  Mismatches   76;  Indels   15;  Gaps    6;

Qy      55 VVGALHGKDCREAVEQLGASANLSEERLAVL--LAGTHTLLQQALRLPPASLKPDAFQEE 112
           :|| ||      |: |: :   |   |||    |  |  :  | |::   :|:|   |::
Db    2347 MVGPLHSSLAASALSQMMSYQGLPSTRLATQPHLVQTQQVQPQNLQMQQQNLQPANIQQQ 2406

Qy     113 --LQ--------ELGIPQDLIGDLASLAFGSQRPLLDSVAQQQGSSLPHVSYFRWRVDVA 162
             ||         ||:     | |   :| |  |    | |  | |   |     :   |
Db    2407 QSLQPPPPPPQPHLGVSSAASGHLGR-SFLSGEPSQADV-QPLGPSSLAVHTILPQESPA 2464

Qy     163 ISTSAQSRSLQPSVLMQLKLTDGSAHRFEVPI 194
           : ||  | || | |     ||  | | :  |:
Db    2465 LPTSLPS-SLVPPVTAAQFLTPPSQHSYSSPV 2495


RESULT  11
US-08-530-950-6
; Sequence 6, Application US/08530950
; Patent No. 5736381
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  Davis, Roger J.
;    APPLICANT:  Raingeaud, Joel
;    APPLICANT:  Gupta, Shashi
;    APPLICANT:  Derijard, Benoit
;    TITLE OF INVENTION:  CYTOKINE-, STRESS-, AND
;    TITLE OF INVENTION:  ONCOPROTEIN-ACTIVATED HUMAN PROTEIN KINASE
;    TITLE OF INVENTION:  KINASES
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/530,950
;      FILING DATE:  19-SEP-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fasse, J. Peter
;      REGISTRATION NUMBER:  32,983
;      REFERENCE/DOCKET NUMBER:  07917/010001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  363 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
US-08-530-950-6


  Query Match             7.1%;  Score 78.5;  DB 1;  Length 363;
  Best Local Similarity   20.9%;  Pred. No. 1.3;
  Matches   57;  Conservative   41;  Mismatches   94;  Indels   81;  Gaps   11;

Qy      20 GRVSFLGSQPSPEVTAV-----------AQLLKDLD---RSTFRKLLKLVVGALHGK-D 63
           | |: :  :|| :: ||            ||| |||   ||:    :    |||  : |
Db      78 GSVNKMVHKPSGQIMAVKRIRSTVDEKEQKQLLMDLDVVMRSSDCPYIVQFYGALFREGD 137

Qy      64 CREAVEQLGASAN-----------------LSEERLAVLLAGTHTLLQQALRLPPASLK 105
           |   :| :  | :                 | :  || : |  |  |:: |::    :|
Db     138 CWICMELMSTSFDKFYKYVYSVLDDVIPEEILGKITLATVKALNH--LKENLKIIHRDIK 195

Qy     106 PD------AFQEELQELGIPQDLIGDLASLAFGSQRPL-----LDSVAQQQGSSL----- 150
           |       :   :| : ||   |:  :|       ||      :|  | :||  :
Db     196 PSNILLDRSGNIKLCDFGISGQLVDSIAKTRDAGCRPYMAPERIDPSASRQGYDVRSDVW 255

Qy     150 -----------PHVSYFRW-----------RVDVAISTSAQSRSLQPSVL--MQLKLTDG 185
                      | :|           : |    :::: |   || :  : | ||
Db     256 SLGITLYELATGRFPYPKWNSVFDQLTQVVKGDPPQLSNSEEREFSPSFINFVNLCLTKD 315

Qy     186 SAHRFEVPIAKFQELRYSVALVLKEMAELEKKC 218
           : |      |::||     ::: |   :|  |
Db     316 ESKR-----PKYKELLKHPFILMYEERAVEVAC 343


RESULT  12
US-08-530-950-10
; Sequence 10, Application US/08530950
; Patent No. 5736381
;  GENERAL INFORMATION:
;    APPLICANT:  Davis, Roger J.
;    APPLICANT:  Raingeaud, Joel
;    APPLICANT:  Gupta, Shashi
;    APPLICANT:  Derijard, Benoit
;    TITLE OF INVENTION:  CYTOKINE-, STRESS-, AND
;    TITLE OF INVENTION:  ONCOPROTEIN-ACTIVATED HUMAN PROTEIN KINASE
;    TITLE OF INVENTION:  KINASES
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/530,950
;      FILING DATE:  19-SEP-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fasse, J. Peter
;      REGISTRATION NUMBER:  32,983
;      REFERENCE/DOCKET NUMBER:  07917/010001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  399 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
US-08-530-950-10


  Query Match             7.1%;  Score 78.5;  DB 1;  Length 399;
  Best Local Similarity   20.9%;  Pred. No. 1.5;
  Matches   57;  Conservative   41;  Mismatches   94;  Indels   81;  Gaps   11;

Qy      20 GRVSFLGSQPSPEVTAV-----------AQLLKDLD---RSTFRKLLKLVVGALHGK-D 63
           | |: :  :|| :: ||            ||| |||   ||:    :    |||  : |
Db     114 GSVNKMVHKPSGQIMAVKRIRSTVDEKEQKQLLMDLDVVMRSSDCPYIVQFYGALFREGD 173

Qy      64 CREAVEQLGASAN-----------------LSEERLAVLLAGTHTLLQQALRLPPASLK 105
           |   :| :  | :                 | :  || : |  |  |:: |::    :|
Db     174 CWICMELMSTSFDKFYKYVYSVLDDVIPEEILGKITLATVKALNH--LKENLKIIHRDIK 231

Qy     106 PD------AFQEELQELGIPQDLIGDLASLAFGSQRPL-----LDSVAQQQGSSL----- 150
           |       :   :| : ||   |:  :|       ||      :|  | :||  :
Db     232 PSNILLDRSGNIKLCDFGISGQLVDSIAKTRDAGCRPYMAPERIDPSASRQGYDVRSDVW 291

Qy     150 -----------PHVSYFRW-----------RVDVAISTSAQSRSLQPSVL--MQLKLTDG 185
                      | :|           : |    :::: |   || :  : | ||
Db     292 SLGITLYELATGRFPYPKWNSVFDQLTQVVKGDPPQLSNSEEREFSPSFINFVNLCLTKD 351

Qy     186 SAHRFEVPIAKFQELRYSVALVLKEMAELEKKC 218
           : |      |::||     ::: |   :|  |
Db     352 ESKR-----PKYKELLKHPFILMYEERAVEVAC 379


RESULT  13
US-08-874-186-92
; Sequence 92, Application US/08874186
; Patent No. 5989885
;  GENERAL INFORMATION:
;    APPLICANT:  Teng, David H-F.
;    APPLICANT:  Tavtigian, Sean V.
;    APPLICANT:  Perry III, William L.
;    APPLICANT:  Skolnick, Mark H.
```

```
;    TITLE OF INVENTION:  SPECIFIC MUTATIONS OF MAP KINASE KINASE
;    TITLE OF INVENTION:  4 (MKK4) IN HUMAN TUMOR CELL LINES IDENTIFY IT AS A TUMOR
;    TITLE OF INVENTION:  SUPPRESSOR IN VARIOUS TYPES OF CANCER
;    NUMBER OF SEQUENCES:  96
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Venable, Baetjer, Howard & Civiletti, LLP
;      STREET:  1201 New York Avenue, N.W., Suite 1000
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/874,186
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/782,482
;      FILING DATE:  10-JAN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Saxe, Stephen A.
;      REGISTRATION NUMBER:  38,609
;      REFERENCE/DOCKET NUMBER:  24884-121392-01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-962-4848
;      TELEFAX:  202-962-8300
;  INFORMATION FOR SEQ ID NO:  92:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  399 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-874-186-92


  Query Match             7.1%;  Score 78.5;  DB 2;  Length 399;
  Best Local Similarity   20.9%;  Pred. No. 1.5;
  Matches   57;  Conservative   41;  Mismatches   94;  Indels   81;  Gaps   11;

Qy      20 GRVSFLGSQPSPEVTAV-----------AQLLKDLD---RSTFRKLLKLVVGALHGK-D 63
           | |: :  :|| :: ||             ||| |||   ||:    :    |||  : |
Db     114 GSVNKMVHKPSGQIMAVKRIRSTVDEKEQKQLLMDLDVVMRSSDCPYIVQFYGALFREGD 173

Qy      64 CREAVEQLGASAN-----------------LSEERLAVLLAGTHTLLQQALRLPPASLK 105
           |   :| :  | :                 | :  || : |  |  |:: |::    :|
Db     174 CWICMELMSTSFDKFYKYVYSVLDDVIPEEILGKITLATVKALNH--LKENLKIIHRDIK 231

Qy     106 PD------AFQEELQELGIPQDLIGDLASLAFGSQRPL-----LDSVAQQQGSSL----- 150
           |       :  :| : ||   |:  :|       ||      :|  | :||  :
Db     232 PSNILLDRSGNIKLCDFGISGQLVDSIAKTRDAGCRPYMAPERIDPSASRQGYDVRSDVW 291

Qy     150 -----------PHVSYFRW-----------RVDVAISTSAQSRSLQPSVL--MQLKLTDG 185
                      | :|           : |    :::: |   || :  : | ||
Db     292 SLGITLYELATGRFPYPKWNSVFDQLTQVVKGDPPQLSNSEEREFSPSFINFVNLCLTKD 351

Qy     186 SAHRFEVPIAKFQELRYSVALVLKEMAELEKKC 218
             : |      |::||     ::: |   :|  |
Db     352 ESKR-----PKYKELLKHPFILMYEERAVEVAC 379


RESULT  14
US-07-642-734C-4
; Sequence 4, Application US/07642734C
; Patent No. 5824513
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, L
;    APPLICANT:  Donadio, S
;    APPLICANT:  Mcalpine, J B
;    TITLE OF INVENTION:  Recombinant DNA Method for Producing
;    TITLE OF INVENTION:  Erythromycin Analogs
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Edward H. Gorman
;      STREET:  Abbott Laboratories D377/AP6D-2 One Abbott
;      STREET:  Park Rd
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  US
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/642,734C
;      FILING DATE:  17-JAN-91
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Danckers, Andreas M
;      REGISTRATION NUMBER:  32652
;      REFERENCE/DOCKET NUMBER:  4952.US.01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  708-937-9396
;      TELEFAX:  708-938-2623
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3567 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-642-734C-4


  Query Match             6.9%;  Score 77;  DB 2;  Length 3567;
  Best Local Similarity   25.3%;  Pred. No. 66;
  Matches   44;  Conservative   23;  Mismatches   45;  Indels   62;  Gaps    9;

Qy      33 VTAVA-----QLLKDL-DRSTFRKLLKLVVGALHGKDCREAVEQLGASANLSEERLAVLL 86
           ||  |     ||:: | ||   | :|    ||  | |   |||: |||        ||:
Db    1136 VTGAASPVGDQLVRWLADRGAERLVL---AGACPGDDLLAAVEEAGAS--------AVVC 1184

Qy      87 AGTHTLLQQALRLPP---------------ASLKPDAFQEELQE----LGIPQDLIGD- 126
           |     |::||   |                 : : |: | | :      | :  :::||
Db    1185 AQDAAALREALGDEPVTALVHAGTLTNFGSISEVAPEEFAETIAAKTALLAVLDEVLGDR 1244

Qy     126 ------------------LASLAFGSQRPLLDSVAQQ---QGSSLPHVSYFRW 157
                            :|: | ||    ||::|:    :| |   |::  |
Db    1245 AVEREVYCSSVAGIWGGAGMAAYAAGS--AYLDALAEHHRARGRSCTSVAWTPW 1296


RESULT  15
US-08-530-950-8
; Sequence 8, Application US/08530950
; Patent No. 5736381
;  GENERAL INFORMATION:
;    APPLICANT:  Davis, Roger J.
;    APPLICANT:  Raingeaud, Joel
;    APPLICANT:  Gupta, Shashi
;    APPLICANT:  Derijard, Benoit
;    TITLE OF INVENTION:  CYTOKINE-, STRESS-, AND
;    TITLE OF INVENTION:  ONCOPROTEIN-ACTIVATED HUMAN PROTEIN KINASE
;    TITLE OF INVENTION:  KINASES
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
```

```
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/530,950
;      FILING DATE:  19-SEP-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fasse, J. Peter
;      REGISTRATION NUMBER:  32,983
;      REFERENCE/DOCKET NUMBER:  07917/010001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  393 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
US-08-530-950-8


  Query Match             6.9%;  Score 76.5;  DB 1;  Length 393;
  Best Local Similarity   21.4%;  Pred. No. 2.5;
  Matches   58;  Conservative   41;  Mismatches   95;  Indels   77;  Gaps   12;

Qy      20 GRVSFLGSQPSPEVTAV-----------AQLLKDLD---RSTFRKLLKLVVGALHGK-D 63
           | |: :  :|| :: ||             ||| |||   ||:    :    |||  : |
Db     108 GSVNKMVHKPSGQIMAVKRIRSTVDEKEQKQLLMDLDVVMRSSDCPYIVQFYGALFREGD 167

Qy      64 CREAVEQLGASAN------------LSEERLA-VLLAGTHTL--LQQALRLPPASLKPD 107
           |   :| :  | :              : || |  : ||    |  |:: |::    :||
Db     168 CWICMELMSTSFDKFYKYVYSVLDDVIPEEILGKITLATVKALMHLKENLKIIHRDIKPS 227

Qy     108 ------AFQEELQELGIPQDLIGDLASLAFGSQRPL-----LDSVAQQQGSSL------- 150
                :   :| : ||   |:  :|       ||      :|  | :||  :
Db     228 NILLDRSGMIKLCDFGISGQLVDSIAKTRDAGCRPYMAPERIDFSASRQGYDVRSDVWSL 287

Qy     150 ---------PHVSYFRW----------RVDVAISTSAQSRSLQPSVL--MQLKLTDGSA 187
                      | :|           : |    :::: |   || :  : | ||   :
Db     288 GITLYELATGRFPYPKWNSVFDQLTQVVKGDPPQLSNSEEREFSPSFINFVNLCLTKDES 347

Qy     188 HRFEVPIAKFQELRYSVALVLKEMAELEKKC 218
           |      |::||     ::: |   :|  |
Db     348 KR-----PKYKELLKHPFILMYEERAVEVAC 373


Search completed: May 15, 2000, 03:22:30
Job time: 5776 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: May 15, 2000, 01:51:10 ; Search time 52.16 Seconds
(without alignments)
251.784 Million cell updates/sec

Title: US-09-223-796-2
Perfect score: 1112
Sequence: 1 MSALGAAAPYLHHPADSHSG.....ALVLKEMAELEKKCERKLQD 224

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 168808 seqs, 58629743 residues

Total number of hits satisfying chosen parameters: 168808

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : PIR_63:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	104.5	9.4	676	2	S41022	hypothetical prote
2	93	8.4	486	2	D72007	CT863 hypothetical
3	91	8.2	584	1	I39710	cellulose biosynth
4	89.5	8.0	433	2	T19632	hypothetical prote
5	89	8.0	3187	2	JC5837	364K Golgi complex
6	84.5	7.6	673	2	T00328	hypothetical prote
7	84.5	7.6	674	2	T19780	hypothetical prote
8	84.5	7.6	1558	2	B71603	RESA-H3 antigen PF
9	84	7.6	303	2	T06981	low-molecular-weig
10	84	7.6	304	2	T06505	glutenin low molec
11	84	7.6	661	2	A55877	tumor necrosis fac
12	84	7.6	790	2	S67803	probable membrane
13	84	7.6	1363	1	A55875	xanthine dehydroge
14	84	7.6	4151	2	T13734	groovin gene prote
15	83.5	7.5	336	2	T26189	hypothetical prote
16	83.5	7.5	1583	2	T14176	probable phosphati
17	83	7.5	200	2	S72612	RNA helicase II -
18	83	7.5	302	2	F69000	cobalamin biosynth
19	82.5	7.4	817	2	T03852	protein phosphatas
20	82.5	7.4	907	2	JG0193	G protein-coupled
21	82	7.4	560	1	WFHUM	mullerian inhibiti
22	81.5	7.3	734	2	E72271	5-methyltetrahydro
23	81.5	7.3	1354	2	T13363	phosphoribosylform
24	81	7.3	459	2	T19347	hypothetical prote
25	81	7.3	1295	2	T24587	hypothetical prote
26	80.5	7.2	315	2	I64065	acetyl-CoA carboxy
27	80.5	7.2	1196	1	GNMVRV	pol polyprotein -
28	80.5	7.2	1325	2	S16129	dynein-associated
29	80.5	7.2	1337	2	A53824	nuclear pore membr
30	80.5	7.2	1479	2	T17401	transcription regu
31	80.5	7.2	2491	2	A57036	talin - slime mold
32	80	7.2	368	2	T04861	hypothetical prote
33	80	7.2	998	2	A45580	transcription fact
34	80	7.2	1048	1	BVECSC	exonuclease (EC 3.
35	80	7.2	1324	2	S52863	DNA-binding protei
36	80	7.2	2555	2	C69681	peptide synthetase
37	80	7.2	3588	2	I40485	surfactin syntheta
38	79.5	7.1	502	2	T35072	probable anthranil
39	79.5	7.1	694	2	C72761	hypothetical prote
40	79.5	7.1	843	1	GNVWK	pol polyprotein -
41	79.5	7.1	1196	1	GNMVGV	pol polyprotein -
42	79.5	7.1	1466	2	S65741	1-phosphatidylinos
43	79.5	7.1	1928	2	JS0610	beta-galactosidase
44	79.5	7.1	2555	2	A40043	notch protein homo
45	79	7.1	431	2	T28697	hypothetical prote

ALIGNMENTS

RESULT 1
S41022
hypothetical protein T07C4.9 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Sep-1997
C;Accession: S41022
R;Berks, M.
submitted to the EMBL Data Library, January 1994
A;Reference number: S41014
A;Accession: S41022
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-676 <BER>
A;Cross-references: EMBL:Z29443; NID:g1067051; PID:g443836
C;Genetics:
A;Introns: 69/3; 161/1; 208/3; 227/1; 357/1; 505/3
C;Superfamily: annexin repeat homology
F;373-444/Domain: annexin repeat homology <AX1>
F;445-516/Domain: annexin repeat homology <AX6>
F;528-600/Domain: annexin repeat homology <AX3>
F;604-675/Domain: annexin repeat homology <AX4>

Query Match 9.4%; Score 104.5; DB 2; Length 676;
Best Local Similarity 25.7%; Pred. No. 0.88;
Matches 47; Conservative 34; Mismatches 53; Indels 49; Gaps 10;

```
Qy      43 LDRSTFRKLLKLVVGALHGKDCREAVEQLGASANLSEERLAVL--LAGTHTLLQQALRLP 100
           |: | |||| :||      |: |       :   ||||:|| |   | | | |
Db      50 LNNSKFRKLCELV------KESR-------SHEMLSEEQLATLYDCATTSTRL------- 90

Qy     101 PASLKPDAFQEELQELGIPQDLI--GDL--ASLAFGS------------QRPLLDSVAQ 143
                |: | :: ::|   :::|  |:|  |  |  |              ::  |: : :
Db      90 -----PNDFADDQEDL---KNIIRYGELFKACHAINSTDFIQKSEDLKDEEKVALERIVE 141

Qy     144 QQ-GSSLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQLKLTDGSAHRFEVPIAKFQELRY 202
           |:   |  :     | |:: :: |  ::||:| : :|:    |  : ||  |  | : |
Db     142 QKLAESAKNEKDIAWNVNIVVANSYVAKSLRPVINIQMPTVGGDTN-FEFDIDSFAQFRQ 200

Qy     203 SVA 205
             :|
Db     201 QLA 203
```

RESULT 2
D72007
CT863 hypothetical protein - Chlamydia pneumoniae (strain CWL029)
C;Species: Chlamydia pneumoniae
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 08-Oct-1999
C;Accession: D72007
R;Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, Nature Genet. 21, 385-389, 1999
A;Title: Comparative genomes of Clamydia pneumoniae and C. trachomatis.

```
A;Reference number: A72000; MUID:99206606
A;Accession: D72007
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-486 <ARN>
A;Cross-references: GB:AE001682; GB:AE001363; NID:g4377344; PIDN:AAD19159.1; PID:g437735
A;Experimental source: strain CWL029
C;Genetics:
A;Gene: CPn1022

  Query Match             8.4%;  Score 93;  DB 2;  Length 486;
  Best Local Similarity   21.9%;  Pred. No. 4.8;
  Matches   54;  Conservative   35;  Mismatches  100;  Indels   58;  Gaps    8;

Qy      27 SQPSPEVTAVAQLLKDLDRSTFRK---LLKLVVG-------ALHGKDCREAVEQLGASAN 76
           : | |     | :    | ||      :||   |        || |    :|:  |
Db       2 NMPVPSAVPSANITLKEDSSTVSTASGILKTATGEVLVSCTALEGSSSTDALISLALGQI 61

Qy      77 LSEERLAVLLAGTHTLLQQALRLPP--------------------ASLKPDAFQ-EELQ 114
           :   :  :||  |:  : | | |||                     : :|   | :
Db      62 ILATQQELLLQSTN--VHQLLFLPPEVVELEIQVVDLLVQLEHAETITSEPQETQTQSRS 119

Qy     115 ELGIPQDLIGDLASLAFGSQRPLLDSVAQQQGSSLPHVSYFR-----------WRVDVAI 163
           |  :||      ::|:  | :| :    |||    |  |  |           |  ::
Db     120 EQTLPQQSSSKQSALSPRSLKPEISDSKQQQALQTPKDSAVRKHSEAPSPETQARASLSQ 179

Qy     164 STSAQSRSLQP-------SVLMQLKLTDGSAHRFEVPIAKFQELRYSVALVLKEMAELEK 216
           ::|:  ||| |        ::| | | :  |      |:::|   :   ||   :  || |
Db     180 ASSSSQRSLPPQESAPERTLLEQQKASSFS------PLSQFSAEKQKEALTTSKSHELYK 233

Qy     217 KCERKLQ 223
           : ::  |
Db     234 ERDQDRQ 240


RESULT   3
I39710
cellulose biosynthesis protein celD - Agrobacterium tumefaciens
C;Species: Agrobacterium tumefaciens
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: I39710
R;Matthysse, A.G.; White, S.; Lightfoot, R.
J. Bacteriol. 177, 1069-1075, 1995
A;Title: Genes required for cellulose synthesis in Agrobacterium tumefaciens.
A;Reference number: I39709; MUID:95164506
A;Accession: I39710
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-584 <RES>
A;Cross-references: GB:L38609; NID:g710486; PIDN:AAC41431.1; PID:g710488
C;Comment: This protein is required for cellulose biosynthesis.
C;Genetics:
A;Gene: celD
C;Superfamily: Agrobacterium tumefaciens cellulose biosynthesis protein celD

  Query Match             8.2%;  Score 91;  DB 1;  Length 584;
  Best Local Similarity   25.0%;  Pred. No. 8.8;
  Matches   53;  Conservative   27;  Mismatches   76;  Indels   56;  Gaps   10;

Qy      22 VSFLGSQPSPEVTAVAQLLKDLDRSTFRKLLKLVVG-----------ALHGKDCREAVE 69
           | : | ||      :  |   :||   : :|  : |             |||   | |:|
Db     263 VLWFGDQP------LGALANIIDRQK-KSVLFYITGRDEDWKTPSPGLVLHGHCIRRAIE 315

Qy      70 Q----------------LG-----ASANLSEERLAVLLAGT------HTLLQQALRLPP 101
           |                |      |  |   |    | ||        : :|||:|
Db     316 QGFKTYDFLRGNEPYKYFFGPEEHKLSCTLFRTRSGDNLGGTLHPRSVRFVYEQALKLYK 375

Qy     102 ASLKPDAFQEELQELGIPQDLIG---DLASLAFGS---QRPLLDSVAQQQGSSLPHVSYF 155
           :  | :|      | |    | :|    ||:|||     ::|      |  | :: |  | |
Db     376 SGKKSEASIAFGQMLSAAPDHLGAQFGLANLAFDRGEFRKPRSPFSASSQQATNP--SSF 433

Qy     156 RWRVDVAI-STSAQSRSLQPSVLMQLKLTDGS 186
               ||:   :   :| ||: :|| :  :::  ||
Db     434 GWRIGETLRCATALSRASEPSQVTNVRVPPGS 465


RESULT   4
T19632
hypothetical protein C32A3.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C;Accession: T19632; T19975
R;Thomas, K.
submitted to the EMBL Data Library, February 1995
A;Reference number: Z19154
A;Accession: T19632
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-433 <WIL>
A;Cross-references: EMBL:Z48241; PIDN:CAA88286.1; GSPDB:GN00021; CESP:C32A3.3
A;Experimental source: clone C32A3
R;Burton, J.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19206
A;Accession: T19975
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-433 <WI2>
A;Cross-references: EMBL:Z81449; PIDN:CAB03766.1; GSPDB:GN00021; CESP:C32A3.3
A;Experimental source: clone C46F11
C;Genetics:
A;Gene: CESP:C32A3.3
A;Map position: 3
A;Introns: 61/3; 94/3; 191/1; 229/3; 298/3; 356/2

  Query Match             8.0%;  Score 89.5;  DB 2;  Length 433;
  Best Local Similarity   23.6%;  Pred. No. 7.9;
  Matches   57;  Conservative   39;  Mismatches   95;  Indels   51;  Gaps   12;

Qy       6 AAAPYLHHPADSHSGRVSFLGSQ------PSPEVTAVAQLLKDLDRSTFRKLLKLVVGAL 59
           :|   || |    |  || : ::      || ::| |   : ||  |      ||:
Db      27 SAVTVLHRPRKKLSAFVSAMAARAASPQSPSKQITVVD--VYDLAASIGNDFEKLIDN- 83

Qy      60 HGKDCREAVEQLGASANLSEERLAVLLAGTHTLLQQALRLPPA--SLKPDAFQEELQELG 117
           :| :|   :     ||    | | : ||      :: :||  |   |: :  |   | |
Db      83 YGNECVRGIMPKVISA---LETLEAMAAGNDRENEEIMRLSKAVERLEQEKHQRNQQHLK 139

Qy     118 IPQDL----------IGDLASLA---FGSQRPLLDSVAQQQGSSLPHVSYFRWRVDVAIS 164
           ::|          | ||  :       | |  :|     ||||:        |  :|
Db     140 FEEELEQVEKTYRKDIDDLQQMVKSLVNENRNLSTTV-----SSLPN------HADSPVS 188

Qy     165 TSAQSRSLQPSVLMQLKLTDGSAHRFEVPI---------AKFQELRYSVALVLKEMAELE 215
           || :   |:  :|::||    |  : |:            : : |: |:  :;:  ||
Db     189 TSMREADLK--MLLELK-EMSSQQKDEIKALQKDVDTYQCQVENLQNSIEKLIRQNEELL 245

Qy     216 KK 217
           :|
Db     246 RK 247


RESULT   5
JC5837
364K Golgi complex-associated protein - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 05-Mar-1998 #sequence_revision 13-Mar-1998 #text_change 20-Aug-1999
C;Accession: JC5837
R;Toki, C.; Fujiwara, T.; Sohda, M.; Hong, H.S.; Misumi, Y.; Ikehara, Y.
Cell Struct. Funct. 22, 565-577, 1997
A;Title: Identification and characterization of rat 364-kDa Golgi-associated protein
A;Reference number: JC5837; MUID:98093490
A;Accession: JC5837
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
```

```
A;Residues: 1-3187 <TOK>
A;Cross-references: DDBJ:D25543; NID:g516825; PIDN:BAA05026.1; PID:d1005567; PID:g516826
C;Comment: This protein plays a role in the formation and maintenance of the characteris
C;Superfamily: giantin
F;49-549,624-1176,1238-1707,1763-3114/Domain: coiled-coil leucine zipper #status predict
F;3165-3187/Domain: membrane anchor #status predicted <MAD>


  Query Match             8.0%;  Score 89;  DB 2;  Length 3187;
  Best Local Similarity   21.8%;  Pred. No. 1.1e+02;
  Matches   49;  Conservative   47;  Mismatches   95;  Indels   34;  Gaps    8;

Qy      11 LHHPADSHSGRVSFLGSQPSPEVTAVAQLLKDLDRSTFRKLLKLVVGALHGKDCREAVEQ 70
           ||    || |  : |  | :  |  :   ||:::    :: |:| : :   :  ::: |
Db    1221 LHAAQPSHPGETATL--QATVSVAQIQDQLKEIE--VEKEELELKISSTTSELTKKSEEV 1276

Qy      71 LGASANLSEERLAVLLAGTHTLLQQALRLPPASLKPDAFQEEL-QELGIPQDLIGDLASL 129
           |     ::|: | :               |  || :  |  |:| |||   |  | ||  |
Db    1277 LLLQEQINEQGLEI-----------QNLKAASHEAKAHTEQLKQELESSQLKIADLEHL 1324

Qy     130 AFGSQRPLLDSVAQQQGSSLPHVSYF------RWRVDVAISTSAQS-----RSLQPSVLM 178
             : :| |::: :  |     |||       : :    : |  :       ::|   : |
Db    1325 K--TLQPELETLQKHVGQKEEEVSYLVGQLGEKEQTLTTVQTEMEEQERLIKALHTQLEM 1382

Qy     179 QLKLTDGSAHRFEVPIAKFQELRYSVALVLKEMAELEKKCERKLQ 223
           | |  :    : :| | : ::       |:| :: ::: :||||
Db    1383 QAKEHEERLKQVQVEICELKKQPKE----LEEESKAKQQLQRKLQ 1423


RESULT   6
T00328
hypothetical protein KIAA0552 - human
C;Species: Homo sapiens (man)
C;Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 08-Oct-1999
C;Accession: T00328
R;Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.
DNA Res. 5, 31-39, 1998
A;Title: Prediction of the coding sequences of unidentified human genes. IX. The complet
A;Reference number: Z14086; MUID:98290545
A;Accession: T00328
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-673 <NAG>
A;Cross-references: EMBL:AB011124; NID:d1185394; PIDN:BAA25478.1; PID:d1026408
A;Experimental source: brain
C;Genetics:
A;Note: KIAA0552


  Query Match             7.6%;  Score 84.5;  DB 2;  Length 673;
  Best Local Similarity   25.2%;  Pred. No. 35;
  Matches   54;  Conservative   34;  Mismatches   67;  Indels   59;  Gaps   12;

Qy      15 ADSHSGRVSFLG-------------SQPSPEVTAVAQLLKDLDRSTFRKLLKLVVGALH 60
           : |  ||   ||               | |||     : |:::|:   : |  :  | ||
Db     287 SSSSMGRPGHLGSGEGGGGGLPFAACSPPSP-----SALIQELEERLWEK--EQEVAAL- 339

Qy      61 GKDCREAVEQLGAS-ANLSEER-------LAVLLAGTHTLLQQALRLPPASLKPDAFQE- 112
             | ::||  |: | : |||        || |  |    |||  |      :    |:
Db     339 ----RRSLEQSEAAVAQVLEERQKAWERELAELRQGCSGKLQQVAR------RAQRAQQG 388

Qy     112 -ELQELGIPQD--LIGDLASLAFGSQRPLLDSVA---QQQGSSLPHVSYFRWRV-----D 160
            :|| | : ||   : : |:      :  | | ||   ::|   || :   :| |     :
Db     389 LQLQVLRLQQDKKQLQEEAARLMRQREELEDKVAACQKEQADFLPRIEETKWEVCQKAGE 448

Qy     161 VAI-------STSAQSRSLQPSVLMQLKLTDGSA 187
           :::       | :  |: |   | :: :| :| |
Db     449 ISLLKQQLKDSQADVSQKLSEIVGLRSQLREGRA 482


RESULT   7
T19780
hypothetical protein C36B1.9 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T19780
R;Lennard, N.
submitted to the EMBL Data Library, September 1996
A;Reference number: Z19176
A;Accession: T19780
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-674 <WIL>
A;Cross-references: EMBL:Z80215; PIDN:CAB02274.1; GSPDB:GN00019; CESP:C36B1.9
A;Experimental source: clone C36B1
C;Genetics:
A;Gene: CESP:C36B1.9
A;Map position: 1
A;Introns: 131/1; 179/1; 266/3; 468/2


  Query Match             7.6%;  Score 84.5;  DB 2;  Length 674;
  Best Local Similarity   21.2%;  Pred. No. 35;
  Matches   39;  Conservative   30;  Mismatches   60;  Indels   55;  Gaps    7;

Qy      63 DCREAVEQLGASANLSEERLAVLLAGTHTLLQQALRLPPASLKPDAFQEELQELG--IPQ 120
           :||:  ||: |:  :: : : :|   |   | : |: ||   :   : |  : ||  ||
Db     111 ECRKINEQVPATVPINFQLMQIL---TTLSLVKVLQTPPQDQEIPKY-ENFERLGTEIP- 166

Qy     121 DLIGDLASLAFGSQRPLLDSVAQQQGSSLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQL 180
             | || :|  |     | :|                 : :       |::|||:
Db     166 --IADLVNLTMGELYEHLKAVFNA---------------IRLRGKRDKTSIRPSVV--- 205

Qy     181 KLTDGSAHRFEVPIAKFQELRYSVALVLKEMA--------------------ELEKKCER 220
                 ||::    |  |:  |:  ||  :                       :|:::  |
Db     205 -------HRYDSEFKKIDEVERSMNRVLHNVIKLQSGETLPRPAPTITIPYWQLDREYRR 257

Qy     221 KLQD 224
           |||
Db     258 SLQD 261


RESULT   8
B71603
RESA-H3 antigen PFB0915w - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 07-May-1999
C;Accession: B71603
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743
A;Accession: B71603
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1558 <GAR>
A;Cross-references: GB:AE001424; GB:AE001362; NID:g3845307; PID:g3845309; TIGR:PFB091
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0915w


  Query Match             7.6%;  Score 84.5;  DB 2;  Length 1558;
  Best Local Similarity   21.6%;  Pred. No. 99;
  Matches   52;  Conservative   45;  Mismatches   79;  Indels   65;  Gaps   10;

Qy      27 SQPSPEVTAVAQLLKDLDRSTFRKLLKLVVGALHGKDCREAVEQLGASANLSEERLAVLL 86
           |  | | |   : :|| ::       : |||  :   |  |:||::    |: ||
Db     951 SSDSKEET---ESIKDKEKD-----VSLVVEEVQDNDMDESVEKVLELKNMEEE------ 997

Qy      87 AGTHTLLQQALRLPPASLKPDAFQEELQELG-IPQDLIGDLASL------------AFG 132
             |:: |: :   : |    || |||  :  ||| |:  |
Db     997 -----LMKDAVEINDITSK--LIEETQELNEVEADLIKDMEKLKELEKALSEDSKEIID 1048

Qy     133 SQRPLLDSVAQQQ---GSSLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQLK-------- 182
```

```
          ::   |: | :::    ::|  |   :   :  |   :  : |:  :| ::|
Db   1049 AKDDTLEKVIEEEHDITTTLDEVVELKDVEEDKIEKVSDLKDLEEDILKEVKEIKELESE 1108

Qy    182 -----------LTDGSAHRFEVP---IAKF----QELRYSVALVLKEMAELEKKCERKLQ 223
                      ||    : |:      ||    :|::   | :|||:: || : |:||:
Db   1109 ILEDYKELKTIETDILEEKKEIEKDHFEKFEEEAEEIKDLEADILKEVSSLEVEEEKKLE 1168

Qy    224 D 224
          :
Db   1169 E 1169


RESULT   9
T06981
low-molecular-weight glutenin storage protein - wheat
C;Species: Triticum aestivum (common wheat)
C;Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 20-Aug-1999
C;Accession: T06981
R;Anderson, O.D.; Cassidy, B.; Dvorak, J.
submitted to the EMBL Data Library, January 1997
A;Reference number: Z15843
A;Accession: T06981
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-303 <AND>
A;Cross-references: EMBL:U86028; NID:g1857655; PIDN:AAB48477.1; PID:g1857656
A;Experimental source: cv. Cheyenne
C;Genetics:
A;Map position: 1
C;Function:
A;Description: seed storage protein in the endosperm
C;Superfamily: gliadin
C;Keywords: seed; storage protein


  Query Match             7.6%;  Score 84;  DB 2;  Length 303;
  Best Local Similarity   27.3%;  Pred. No. 14;
  Matches   38;  Conservative   18;  Mismatches   67;  Indels   16;  Gaps    4;

Qy     99 LPPA---SLKPDAFQEELQELGIPQDLIGDLASLAFGSQRPLLDS---VAQQQGSSLPHV 152
          |||    | :|   |:: | |  ||          |  |:|:|      :|||   ||
Db     38 LPPQQSFSQQPPFSQQQQQPL--PQQPSFSQQQPPFSQQQPILSQQPPFSQQQQPVLPQQ 95

Qy    153 SYFRWRVDVAISTSAQSRSL--------QPSVLMQLKLTDGSAHRFEVPIAKFQELRYSV 204
          | |  :  : :    | : |        ||||| ||        :   |:|  | |  |
Db     96 SPFSQQQQLVLPPQQQQQQLVQQQIPIVQPSVLQQLNPCKVFLQQQCSPVAMPQRLARSQ 155

Qy    205 ALVLKEMAELEKKCERKLQ 223
                :::::| ::||
Db    156 MWQQSSCHVMQQQCCQQLQ 174


RESULT  10
T06505
glutenin low molecular weight chain precursor (B-I) - wheat
C;Species: Triticum aestivum (common wheat)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 12-Nov-1999
C;Accession: T06505
R;Okita, T.W.; Cheesbrough, V.; Reeves, C.D.
J. Biol. Chem. 260, 8203-8213, 1985
A;Title: Evolution and heterogeneity of the alpha/beta-type and gamma-type gliadin DNA s
A;Reference number: A92541; MUID:85234522
A;Accession: T06505
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-304 <OKI>
A;Cross-references: EMBL:M11077; NID:g170729; PIDN:AAA34285.1; PID:g170730
C;Superfamily: gliadin
C;Keywords: seed; storage protein
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-304/Product: glutenin low molecular weight chain (B-I) #status predicted <MAT>


  Query Match             7.6%;  Score 84;  DB 2;  Length 304;
  Best Local Similarity   27.3%;  Pred. No. 14;
  Matches   38;  Conservative   18;  Mismatches   67;  Indels   16;  Gaps    4;

Qy     99 LPPA---SLKPDAFQEELQELGIPQDLIGDLASLAFGSQRPLLDS---VAQQQGSSLPHV 152
          |||    | :|   |:: | |  ||          |  |:|:|      :|||   ||
Db     38 LPPQQSFSQQPPFSQQQQQPL--PQQPSFSQQQPPFSQQQPILSQQPPFSQQQQPVLPQQ 95

Qy    153 SYFRWRVDVAISTSAQSRSL--------QPSVLMQLKLTDGSAHRFEVPIAKFQELRYSV 204
          | |  :  : :    | : |        ||||| ||        :   |:|  | |  |
Db     96 SPFSQQQQLVLPPQQQQQQLVQQQIPIVQPSVLQQLNPCKVFLQQQCSPVAMPQRLARSQ 155

Qy    205 ALVLKEMAELEKKCERKLQ 223
                :::::| ::||
Db    156 MWQQSSCHVMQQQCCQQLQ 174


RESULT  11
A55877
tumor necrosis factor type 1 receptor associated protein TRAP-1 - human
C;Species: Homo sapiens (man)
C;Date: 23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change 05-Nov-1999
C;Accession: A55877
R;Song, H.Y.; Dunbar, J.D.; Zhang, Y.X.; Guo, D.; Donner, D.B.
J. Biol. Chem. 270, 3574-3581, 1995
A;Title: Identification of a protein with homology to hsp90 that binds the type 1 tum
A;Reference number: A55877; MUID:95181307
A;Accession: A55877
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-661 <SON>
A;Cross-references: GB:U12595; NID:g687236; PIDN:AAA87704.1; PID:g687237
C;Superfamily: heat shock protein 90


  Query Match             7.6%;  Score 84;  DB 2;  Length 661;
  Best Local Similarity   23.8%;  Pred. No. 37;
  Matches   46;  Conservative   39;  Mismatches   76;  Indels   32;  Gaps   10;

Qy     49 RKLLKLVVGALHGKDCREA-VEQLGASANLSEERLAVLLAGTHTLLQQALRLPPASLKPD 107
          :||| :|  :|: :  :|  : :| ::|: : |:|       | |:     ||   :
Db     80 KKLLDIVARSLYSE--KEVFIRELISNASDALEKL------RHKLVSDGQALPEMEIHLQ 131

Qy    108 AFQEE----LQELGI---PQDLIGDLASLAFGSQRPLLDSVAQQQGSSLPHVSY----FR 156
             |:    :|: ||    ::|: :| ::|    :  ||::  |  :|  :     |
Db    132 TNAEKGTITIQDTGIGMTQEELVSNLGTIARSGSKAFLDALQNQAEASSKIIGQFGVGFY 191

Qy    157 WRVDVAISTSAQSRSLQPSVLMQLKLTDGSAHRFEVPIAKFQELRYSVALVLKEMAELEK 216
              ||      |||  |  |    |:|||    ||  ||:   :|    :::     |:
Db    192 SAFMVADRVEVYSRSAAPGSLGYQWLSDGSG-VFE--IAEASGVRTGTKIII---HLKS 244

Qy    217 KC-----ERKLQD 224
            |     | :::|
Db    245 DCKEFSSEARVRD 257


RESULT  12
S67803
probable membrane protein YDL239c - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein D0771
C;Species: Saccharomyces cerevisiae
C;Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 29-Oct-1999
C;Accession: S67803
R;Alt-Moerbe, J.; Schneider, C.; Moro, M.
submitted to the Protein Sequence Database, July 1996
A;Reference number: S67798
A;Accession: S67803
A;Molecule type: DNA
A;Residues: 1-790 <ALT>
A;Cross-references: EMBL:Z74287; NID:g1431405; PID:e253369; PID:g1431406; GSPDB:GN000
A;Experimental source: strain S288C
C;Genetics:
A;Gene: MIPS:YDL239c
```

A;Map position: 4L
C;Keywords: transmembrane protein
F;706-722/Domain: transmembrane #status predicted <TMM>

Query Match 7.6%; Score 84; DB 2; Length 790;
Best Local Similarity 21.4%; Pred. No. 46;
Matches 57; Conservative 41; Mismatches 86; Indels 82; Gaps 12;

```
Qy     11 LHHPADSHSGRVSFLGSQPSPEVTAVAQL---------LKDLDRSTFRKLLKLVVGAL-- 60
Db    517 LHFMQQQHSHKIKYLQKE---ALTKEQQFRLEKRRWHDILNLKEENFQKLKSELKGKLIL 573

Qy     60 --------------HGKDCREAVEQLGASANL-------------SEERLAVLLAGTHTL 92
Db    574 SEKIQKNAEDKLNDYMNEHQEIVEKLQNQALIASRWSTQIQESENTHKKITDELAGKQS- 633

Qy     93 LQQALRLPPA--SLKPDAFQEELQELGIPQDLIGDLASLAFGSQRPLLDSVAQQQGSSLP 150
Db    633 --EILKLEETILSLKEDVFQEKLN----LKKLYGD----------PSTELNFETVGKSFP 676

Qy    151 HVS---YFRWRVDVAIS-TSAQSRSLQPSVLMQLKLTDGSAHRFEVPIAKFQ-------- 199
Db    677 HITKEKYDSLGLDILTDLTYVQSQNLIKNLLIVL----------DIPLKTFLKIVPTIVI 726

Qy    199 ELRYSVALVLKEMAELEKKCERKLQD 224
Db    727 QLRCELTLLTKFANDLNLKVFGKQLD 752
```

RESULT 13
A55875
xanthine dehydrogenase (EC 1.1.1.204) - Emericella nidulans
N;Alternate names: purine hydroxylase I
C;Species: Emericella nidulans, Aspergillus nidulans
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A55875; S49819
R;Glatigny, A.; Scazzocchio, C.
J. Biol. Chem. 270, 3534-3550, 1995
A;Title: Cloning and molecular characterization of hxA, the gene coding for the xanthine
A;Reference number: A55875; MUID:95181302
A;Accession: A55875
A;Molecule type: DNA
A;Residues: 1-1363 <GLA>
A;Cross-references: EMBL:X82827; NID:g577730; PIDN:CAA58034.1; PID:g577731
C;Genetics:
A;Gene: hxA
A;Introns: 83/3; 151/3; 803/2
C;Superfamily: xanthine dehydrogenase; ferredoxin [2Fe-2S] homology
C;Keywords: 2Fe-2S; FAD; flavoprotein; iron-sulfur protein; metalloprotein; molybdenum;
F;57-104/Domain: ferredoxin [2Fe-2S] homology <FER1>
F;73,78,81,103/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted

Query Match 7.6%; Score 84; DB 1; Length 1363;
Best Local Similarity 22.5%; Pred. No. 92;
Matches 46; Conservative 35; Mismatches 81; Indels 42; Gaps 10;

```
Qy     23 SFLGSQPSPEVTAVAQLLK-DLDRST-FRKLLKLVVGALHGKDCREAVEQLGASANLSEE 80
Db    155 ALLRNDPKPSEHAVEEAFDGNLCRCTGYRPILDAAQSFTSPIGCGKARANGGSGCCMEEQ 214

Qy     81 RLAVLLAGTHTLLQQALRLPPASLK-----PDAFQEELQ-ELGIPQDL-IGDLASLAFGS 133
Db    215 K------GTNGCCKGSSEETTEDVKHKFASPDFIEYKPDTELIFPPSLWKHELRPLAFGN 268

Qy    134 QRPLLDSVAQQQGSSLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQLKLTDGSAH-RFEV 192
Db    269 KRK------------------KWYRPVTVQQLLEIKSIHPDA----KLIGGSTETQIEI 305

Qy    193 PIAKFQELRYSVALVLKEMAELEK 216
Db    306 ---KFKQMRYGASVYLGDLAELRQ 326
```

RESULT 14
T13734
groovin gene protein - fruit fly (Drosophila melanogaster) (fragment)
C;Species: Drosophila melanogaster
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 13-Aug-1999
C;Accession: T13734
R;Strumpf, D.; Volk, T.
J. Cell Biol. 143, 1259-1270, 1998
A;Title: Kakapo, a novel cytoskeletal-associated protein is essential for the restric
A;Reference number: Z17746
A;Accession: T13734
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-4151 <STR>
A;Cross-references: EMBL:Y09430; NID:e1372000; PID:e1372001; PIDN:CAA70581.1
C;Genetics:
A;Gene: grv
A;Map position: 2

Query Match 7.6%; Score 84; DB 2; Length 4151;
Best Local Similarity 19.1%; Pred. No. 3.7e+02;
Matches 56; Conservative 51; Mismatches 108; Indels 78; Gaps 10;

```
Qy      1 MSALGAAAPYLHHPADSHSGRVSFLG---------SQPS---PEVT-------------- 35
Db   2430 LGILEQALPLSEHFADSHQGLTAWLDDMEQQISRLSMPALRPDQITLQQDKNERLLQSIA 2489

Qy     35 --------------AVAQLLKDLDRSTFRKLL-----KLVVGALHGKDCREAVEQLGASA 75
Db   2490 EHKPLLDKLNKTGEALGALVADDDGAKINEILDTDNARYAALRLELRERQQALESALQES 2549

Qy     76 NLSEERLAVLLAGTHTLLQQALRLPPASLKPDAFQEELQELGIPQDLIGDLASLAFGS-Q 134
Db   2550 SQFSDKLEGMLRALANTVDQVNQLDPLSALPQKIREQIEDNDALMDDL-DKRQDAFSAVQ 2608

Qy    135 RPLLDSVAQQQGSSLPHVSYFRWRV--------DVAISTSAQSRSL-------QP----- 175
Db   2609 RAANDVIAKAGNKADPAVRDVKAKLEKLNNLWNDVQNATKKRGSSLDDILSVAEPFWKQL 2668

Qy    175 -SVLMQLKLTDGSAHRFEVPIAK----------FQELRYSVALVLKEMAELEK 216
Db   2669 NSVMKTLKDLEETLSCQEPPAAQPQDIKKQQVALQEIRHEIDQTKPEVEQVRR 2721
```

RESULT 15
T26189
hypothetical protein W05B5.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T26189
R;McLay, K.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z20167
A;Accession: T26189
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-336 <WIL>
A;Cross-references: EMBL:Z82071; NID:e1298254; PIDN:CAB04919.1; GSPDB:GN00019; CESP:W
A;Experimental source: clone W05B5
C;Genetics:
A;Gene: CESP:W05B5.3
A;Map position: 1
A;Introns: 5/1; 59/3; 109/3; 233/3

Query Match 7.5%; Score 83.5; DB 2; Length 336;
Best Local Similarity 24.8%; Pred. No. 17;
Matches 51; Conservative 24; Mismatches 78; Indels 53; Gaps 8;

```
Qy      6 AAAPYLHHPADSHSGRVSFLGSQPSPEVTAVAQLLKDLDRSTFRKLLK-----LVVGALH 60
```

```
          :| | :     | | |:  |  |  | :||:|    ||   : : :|       |:| |
Db    140 SANPIIPQAIQSISARIRQLPPQTVPILTAIAVCQADLLPESQQPMLLAERLWCVLGKLG 199

Qy     61 GKDCREAVEQLGASANLSEERLAVLLAGTHTLLQ---QALR------------------ 99
          |      :: |  : :        |||   || |     ||
Db    200 G------IQSLATAPS--------LLADVRTLRQWHSDRLRQMSQISQHFSQNLLIAPVA 245

Qy     99 ------LPPASLKPDAFQEELQELGIPQDLIGDLASLAFGSQRP--LLDSVAQQQ-GSSL 149
                |||| | | |         :  : |    |:|   :||  :  | ||:    ||
Db    246 AAAPVLLPPAFLSPPASATSTSSSSVKSEFIERHPSIASLLERPRRISSSGAQEPLNLSL 305

Qy    150 PHVSYFRWRVDVAISTSAQSRSLQPS 175
          |||   | :|   : :  |   :: |
Db    306 PHV---RHQVKRDVDSDEQLEEMKVS 328


Search completed: May 15, 2000, 03:23:35
Job time: 5545 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: May 15, 2000, 02:06:10 ; Search time 44.1 Seconds
(without alignments)
154.692 Million cell updates/sec

Title: US-09-223-796-2
Perfect score: 1112
Sequence: 1 MSALGAAAPYLHHPADSHSG.....ALVLKEMAELEKKCERKLQD 224

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 83857 seqs, 30454973 residues

Total number of hits satisfying chosen parameters: 83857

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : SwissProt_38:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	84.5	7.6	673	1	Y552_HUMAN	O60299 homo sapien
2	84.5	7.6	890	1	RB6K_HUMAN	O95235 homo sapien
3	84	7.6	304	1	GDB1_WHEAT	P04729 triticum ae
4	84	7.6	661	1	TRAL_HUMAN	Q12931 homo sapien
5	84	7.6	1363	1	XDH_EMENI	Q12553 emericella
6	83	7.5	302	1	COBK_METTH	O27083 methanobact
7	82	7.4	560	1	MIS_HUMAN	P03971 homo sapien
8	81.5	7.3	1354	1	PUR4_DROME	P35421 drosophila
9	80.5	7.2	315	1	ACCA_HAEIN	P43872 haemophilus
10	80.5	7.2	1196	1	POL_MLVRD	P11227 radiation m
11	80.5	7.2	1280	1	DYNA_RAT	P28023 rattus norv
12	80.5	7.2	1337	1	P152_YEAST	P39685 saccharomyc
13	80.5	7.2	2491	1	TALA_DICDI	P54633 dictyosteli
14	80	7.2	314	1	IUNH_CRIFA	Q27546 crithidia f
15	80	7.2	887	1	RB6K_MOUSE	P97329 mus musculu
16	80	7.2	1048	1	SBCC_ECOLI	P13458 escherichia
17	80	7.2	2555	1	PPS3_BACSU	P39847 bacillus su
18	79.5	7.1	572	1	PRIM_STAAU	O05338 staphylococ
19	79.5	7.1	843	1	POL_MLVAK	P03357 akr murine
20	79.5	7.1	1196	1	POL_MLVAV	P03356 akv murine
21	79.5	7.1	1928	1	LPH_RAT	Q02401 rattus norv
22	79	7.1	1194	1	BCHH_RHOCA	P26162 rhodobacter
23	79	7.1	1270	1	DYNA_HUMAN	Q14203 homo sapien
24	78.5	7.1	397	1	MPK4_MOUSE	P47809 m dual spec
25	78.5	7.1	399	1	MPK4_HUMAN	P45985 homo sapien
26	78.5	7.1	791	1	TEX_BORPE	Q45388 bordetella
27	78.5	7.1	1281	1	DYNA_MOUSE	O08788 mus musculu
28	78.5	7.1	1427	1	REST_HUMAN	P30622 homo sapien
29	78	7.0	405	1	FLIK_SALTY	P26416 salmonella
30	78	7.0	1959	1	MYSN_CHICK	P14105 gallus gall
31	78	7.0	3587	1	SRF1_BACSU	P27206 bacillus su
32	77.5	7.0	487	1	AMPL_BOVIN	P00727 bos taurus
33	77.5	7.0	799	1	CAD8_MOUSE	P97291 mus musculu
34	77.5	7.0	962	1	IF3A_MAIZE	Q9xhr2 zea mays (m
35	77	6.9	164	1	GRP_ZYMMO	P74996 zymomonas m
36	77	6.9	689	1	ARK1_BOVIN	P21146 bos taurus
37	77	6.9	768	1	TRPG_EMENI	P06531 emericella
38	77	6.9	1947	1	MYSC_CAEEL	P12845 caenorhabdi
39	77	6.9	3567	1	ERY2_SACER	Q03132 saccharopol
40	76.5	6.9	757	1	YHGF_NEIME	Q51152 neisseria m
41	76.5	6.9	1204	1	POL_MLVFF	P26809 friend muri
42	76.5	6.9	1301	1	DDX9_CAEEL	Q22307 caenorhabdi
43	76	6.8	395	1	METC_BORAV	Q07703 bordetella
44	76	6.8	689	1	ARK1_HUMAN	P25098 homo sapien
45	76	6.8	1231	1	YKT3_CAEEL	P34314 caenorhabdi

ALIGNMENTS

RESULT 1
Y552_HUMAN
ID Y552_HUMAN STANDARD; PRT; 673 AA.
AC O60299;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL PROTEIN KIAA0552.
GN KIAA0552.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE; 98290545.
RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
RA Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:31-39(1998).
CC -!- SIMILARITY: TO HUMAN KIAA0341.
CC ---

CC ---
DR EMBL; AB011124; BAA25478.1; -.
KW Hypothetical protein; Coiled coil.
FT DOMAIN 317 496 COILED COIL (POTENTIAL).
FT DOMAIN 571 639 COILED COIL (POTENTIAL).
FT DOMAIN 60 63 POLY-GLY.
FT DOMAIN 122 125 POLY-SER.
FT DOMAIN 282 290 POLY-SER.
FT DOMAIN 298 306 POLY-GLY.
FT DOMAIN 550 553 POLY-ALA.
SQ SEQUENCE 673 AA; 71790 MW; 3E548EC03A01F770 CRC64;

Query Match 7.6%; Score 84.5; DB 1; Length 673;
Best Local Similarity 25.2%; Pred. No. 16;
Matches 54; Conservative 34; Mismatches 67; Indels 59; Gaps 12;

```
Qy     15 ADSHSGRVSFLG----------SQPSPEVTAVAQLLKDLDRSTFRKLLKLVVGALH 60
          : |  ||  ||             | |||     : |:::|:   : |  :  | ||
Db    287 SSSSMGRPGHLGSGEGGGGGLPFAACSPPSP-----SALIQELEERLWEK--EQEVAAL- 339

Qy     61 GKDCREAVEQLGAS-ANLSEER-------LAVLLAGTHTLLQQALRLPPASLKPDAFQE- 112
               | ::||  |: | : |||        || |  |    |||  |      :    |:
Db    339 ----RRSLEQSEAAVAQVLEERQKAWERELAELRQGCSGKLQQVAR------RAQRAQQG 388

Qy    112 -ELQELGIPQD--LIGDLASLAFGSQRPLLDSVA---QQQGSSLPHVSYFRWRV-----D 160
```

```
          :|| | : ||   : : |:     :  | | ||   ::|   || :   :| |     :
Db    389 LQLQVLRLQQDKKQLQEEAARLMRQREELEDKVAACQKEQADFLPRIEETKWEVCQKAGE 448

Qy    161 VAI-------STSAQSRSLQPSVLMQLKLTDGSA 187
          :::       | :  |: |   | :: :| :| |
Db    449 ISLLKQQLKDSQADVSQKLSEIVGLRSQLREGRA 482


RESULT   2
RB6K_HUMAN
ID   RB6K_HUMAN     STANDARD;      PRT;   890 AA.
AC   O95235;
DT   15-FEB-2000 (Rel. 39, Created)
DT   15-FEB-2000 (Rel. 39, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   RABKINESIN-6 (RAB6-INTERACTING KINESIN-LIKE PROTEIN) (GG10_2).
GN   RAB6KIFL.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=ENDOTHELIAL CELLS;
RX   MEDLINE; 99252096.
RA   Horrevoets A.J.G., Fontijn R.D., van Zonneveld A.J., de Vries C.J.M.,
RA   ten Cate J.W., Pannekoek H.;
RT   "Vascular endothelial genes that are responsive to tumor necrosis
RT   factor-alpha in vitro are expressed in atherosclerotic lesions,
RT   including inhibitor of apoptosis protein-1, stannin, and two novel
RT   genes.";
RL   Blood 93:3418-3431(1999).
CC   -!- FUNCTION: INTERACTS WITH GUANOSINE TRIPHOSPHATE (GTP)-BOUND FORMS
CC       OF RAB6. MAY ACT AS A MOTOR REQUIRED FOR THE RETROGRADE RAB-6
CC       REGULATED TRANSPORT OF GOLGI MEMBRANES AND ASSOCIATED VESICLES
CC       ALONG MICROTUBLES. HAS A MICROTUBULE PLUS END-DIRECTED MOTILITY
CC       (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: GOLGI (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF070672; AAC83230.1; -.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; FALSE_NEG.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil; Golgi stack;
KW   Protein transport.
FT   DOMAIN       70    495       MECHANOCHEMICAL (MOTOR).
FT   NP_BIND     160    167       ATP (POTENTIAL).
FT   DOMAIN      611    762       COILED COIL (POTENTIAL).
FT   DOMAIN      763    890       GLOBULAR (POTENTIAL).
SQ   SEQUENCE   890 AA;  100277 MW;  6620264615496051 CRC64;


  Query Match             7.6%;  Score 84.5;  DB 1;  Length 890;
  Best Local Similarity   24.5%;  Pred. No. 22;
  Matches   64;  Conservative   32;  Mismatches   98;  Indels   67;  Gaps   13;

Qy      9 PYLHHPADSHSGRVSFLGSQPSPEVTAVA------QLLKDLDRSTFRK--LLKLVVG--- 58
          | ||     || :||     || |  | |       :  : | | : |  ||::|
Db    518 PSLHSFIKEHSLQVS-----PSLEKGAKADTGLDDDIENEADISMYGKEELLQVVEAMKT 572

Qy     58 ------------ALHGKD--CREAVEQLGASANLSEERLAVLLAGTHTLLQQALRLPPAS 103
                      :| :|  | | |||:        | |         : :: | :   |
Db    573 LLLKERQEKLQLEMHLRDEICNEMVEQMQQREQWCSEHLDTQKELLEEMYEEKLNILKES 632

Qy    104 LKPDAFQEELQELGIPQDLIGDLASLAFGSQRPLLDSVA-QQQGSSLPHVSYFRWRVDVA 162
          |    :|||:||     : | :| :|    |     ||| || || |       |    :|
Db    633 L-TSFYQEEIQE---RDEKIEELEALL---QEARQQSVAHQQSGSELA----LRRSQRLA 681

Qy    163 ISTSAQSRSLQPSVLMQLKL----TDGSAHRFEVPI-------------AKFQELRYSV 204
          | | |      : | | |      |   |::: :                 | :| : ::
Db    682 ASASTQQLQEVKAKLQQCKALNSTTEELHKYQKMLEPPPSAKPFTIDVDKKLEEGQKNI 741

Qy    205 ALVLKEMAEL-------EKKC 218
          |:  |: :|        |: |
Db    742 RLLRTELQKLGESLQSAERAC 762


RESULT   3
GDB1_WHEAT
ID   GDB1_WHEAT     STANDARD;      PRT;   304 AA.
AC   P04729;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   GAMMA-GLIADIN B-I PRECURSOR.
OS   Triticum aestivum (Wheat).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
OC   Poaceae; Triticum.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 85234522.
RA   Okita T.W., Cheesbrough V., Reeves C.D.;
RT   "Evolution and heterogeneity of the alpha-/beta-type and gamma-type
RT   gliadin DNA sequences.";
RL   J. Biol. Chem. 260:8203-8213(1985).
CC   -!- FUNCTION: GLIADIN IS THE MAJOR SEED STORAGE PROTEIN IN WHEAT.
CC   -!- MISCELLANEOUS: THE GAMMA-GLIADINS CAN BE DIVIDED INTO 3 HOMOLOGY
CC       CLASSES. SEQUENCE DIVERGENCE BETWEEN THE CLASSES IS DUE TO
CC       SINGLE-BASE SUBSTITUTIONS & TO DUPLICATIONS OR DELETIONS WITHIN OR
CC       NEAR DIRECT REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M11077; AAA34285.1; -.
DR   PRINTS; PR00208; GLIADGLUTEN.
KW   Seed storage protein; Repeat; Signal; Multigene family.
FT   SIGNAL       1     23
FT   CHAIN       24    304       GAMMA-GLIADIN B-I.
SQ   SEQUENCE   304 AA;  34252 MW;  807EBF447A59D6D5 CRC64;


  Query Match             7.6%;  Score 84;  DB 1;  Length 304;
  Best Local Similarity   27.3%;  Pred. No. 6.4;
  Matches   38;  Conservative   18;  Mismatches   67;  Indels   16;  Gaps    4;

Qy     99 LPPA---SLKPDAFQEELQELGIPQDLIGDLASLAFGSQRPLLDS---VAQQQGSSLPHV 152
          |||    | :|   |:: | |  ||          |  |:|:|      :|||   ||
Db     38 LPPQQSFSQQPPFSQQQQQPL--PQQPSFSQQQPPFSQQQPILSQQPPFSQQQQPVLPQQ 95

Qy    153 SYFRWRVDVAISTSAQSRSL--------QPSVLMQLKLTDGSAHRFEVPIAKFQELRYSV 204
          | |  :  : :    | : |        |||| ||        :   |:|  | |  |
Db     96 SPFSQQQQLVLPPQQQQQQLVQQQIPIVQPSVLQQLNPCKVFLQQQCSPVAMPQRLARSQ 155

Qy    205 ALVLKEMAELEKKCERKLQ 223
                 ::::| ::||
Db    156 MWQQSSCHVMQQQCCQQLQ 174


RESULT   4
TRAL_HUMAN
ID   TRAL_HUMAN     STANDARD;      PRT;   661 AA.
AC   Q12931; O75235;
```

```
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   TUMOR NECROSIS FACTOR TYPE 1 RECEPTOR ASSOCIATED PROTEIN (TRAP-1)
DE   (FRAGMENT).
GN   TRAP1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95181307.
RA   Song H.Y., Dunbar J.D., Zhang Y.X., Guo D., Donner D.B.;
RT   "Identification of a protein with homology to hsp90 that binds the
RT   type 1 tumor necrosis factor receptor.";
RL   J. Biol. Chem. 270:3574-3581(1995).
RN   [2]
RP   SEQUENCE OF 16-631 FROM N.A.
RA   Ricke D.O., Bruce D., Mundt M., Doggett N., Munk C., Saunders E.,
RA   Robinson D., Jones M., Buckingham J., Chasteen L., Thompson S.,
RA   Goodwin L., Bryant J., Tesmer J., Meincke L., Longmire J., White S.,
RA   Ueng S., Tatum O., Campbell C., Fawcett J., Maltbie M., Misra M.,
RA   Deaven L.;
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: BINDS TO THE INTRACELLULAR DOMAIN OF TUMOR NECROSIS
CC       FACTOR TYPE 1 RECEPTOR.
CC   -!- TISSUE SPECIFICITY: FOUND IN SKELETAL MUSCLE, LIVER, HEART, BRAIN,
CC       KINDEY, PANCREAS, LUNG AND PLACENTA.
CC   -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U12595; AAA87704.1; -.
DR   EMBL; AC005203; AAC24722.1; -.
DR   PROSITE; PS00298; HSP90; FALSE_NEG.
DR   PFAM; PF00183; HSP90; 2.
FT   NON_TER       1      1
FT   CONFLICT    292    292       G -> R (IN REF. 2).
SQ   SEQUENCE   661 AA;  75342 MW;  EF527B93965032E5 CRC64;


  Query Match             7.6%;  Score 84;  DB 1;  Length 661;
  Best Local Similarity   23.8%;  Pred. No. 17;
  Matches   46;  Conservative   39;  Mismatches   76;  Indels   32;  Gaps   10;

Qy      49 RKLLKLVVGALHGKDCREA-VEQLGASANLSEERLAVLLAGTHTLLQQALRLPPASLKPD 107
           :||| :|  :|: :  :|  : :| ::|: : |:|       | |:     ||   :
Db      80 KKLLDIVARSLYSE--KEVFIRELISNASDALEKL------RHKLVSDGQALPEMEIHLQ 131

Qy     108 AFQEE----LQELGI---PQDLIGDLASLAFGSQRPLLDSVAQQQGSSLPHVSY----FR 156
             |:    :|: ||    ::|: :| ::|    :  ||::  |  :|  :     |
Db     132 TNAEKGTITIQDTGIGMTQEELVSNLGTIARSGSKAFLDALQNQAEASSKIIGQFGVGFY 191

Qy     157 WRVDVAISTSAQSRSLQPSVLMQLKLTDGSAHRFEVPIAKFQELRYSVALVLKEMAELEK 216
           '     ||       |||  |  |    |:|||   ||  ||:   :|    :::      |:
Db     192 SAFMVADRVEVYSRSAAPGSLGYQWLSDGSG-VFE--IAEASGVRTGTKIII----HLKS 244

Qy     217 KC-----ERKLQD 224
           |     | :::|
Db     245 DCKEFSSEARVRD 257


RESULT   5
XDH_EMENI
ID   XDH_EMENI      STANDARD;      PRT;  1363 AA.
AC   Q12553;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   XANTHINE DEHYDROGENASE (EC 1.1.1.204) (PURINE HYDROXYLASE I).
GN   HXA.
OS   Emericella nidulans (Aspergillus nidulans).
OC   Eukaryota; Fungi; Ascomycota; Plectomycetes; Eurotiales;
OC   Trichocomaceae; Emericella.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BIA1;
RX   MEDLINE; 95181302.
RA   Glatigny A., Scazzocchio C.;
RT   "Cloning and molecular characterization of hxA, the gene coding for
RT   the xanthine dehydrogenase (purine hydroxylase I) of Aspergillus
RT   nidulans.";
RL   J. Biol. Chem. 270:3534-3550(1995).
CC   -!- CATALYTIC ACTIVITY: XANTHINE + NAD(+) + H(2)O = URATE + NADH.
CC   -!- COFACTOR: FAD, MOLYBDOPTERIN, AND TWO 2FE-2S CLUSTERS.
CC   -!- SUBCELLULAR LOCATION: PEROXISOMAL.
CC   -!- SIMILARITY: TO OTHER XANTHINE DEHYDROGENASES, AND LIMITED TO
CC       OTHER EUKARYOTIC MOLYBDOPTERIN ENZYMES SUCH AS NITRATE REDUCTASE
CC       AND SULFITE OXIDASE.
CC   -!- SIMILARITY: TO 2FE-2S FERREDOXINS IN THE N-TERMINAL DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X82827; CAA58034.1; -.
DR   PROSITE; PS00197; 2FE2S_FERREDOXIN; 1.
DR   PROSITE; PS00559; MOLYBDOPTERIN_EUK; FALSE_NEG.
DR   PFAM; PF00941; dehydrog_molyb; 1.
DR   PFAM; PF01315; Ald_Xan_dh_C; 1.
KW   Oxidoreductase; NAD; Molybdenum; Flavoprotein; FAD; Iron-sulfur;
KW   Peroxisome.
FT   METAL        73     73       IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT   METAL        78     78       IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT   METAL        81     81       IRON-SULFUR (2FE-2S) (BY SIMILARITY).
SQ   SEQUENCE   1363 AA;  149504 MW;  F9464E66C22ACD97 CRC64;


  Query Match             7.6%;  Score 84;  DB 1;  Length 1363;
  Best Local Similarity   22.5%;  Pred. No. 41;
  Matches   46;  Conservative   35;  Mismatches   81;  Indels   42;  Gaps   10;

Qy      23 SFLGSQPSPEVTAVAQLLK-DLDRST-FRKLLKLVVGALHGKDCREAVEQLGASANLSEE 80
           : | : | |   || :   | | | :| :|          | :|    |:   : |:
Db     155 ALLRNDPKPSEHAVEEAFDGNLCRCTGYRPILDAAQSFTSPIGCGKARANGGSGCCMEEQ 214

Qy      81 RLAVLLAGTHTLLQQALRLPPASLK-----PDAFQEELQ-ELGIPQDL-IGDLASLAFGS 133
           :      ||:   : :       :|     ||  : :   ||  |  |   :|  ||||:
Db     215 K------GTNGCCKGSSEETGEDVKHKFASPDFIEYKPDTELIFPPSLWKHELRPLAFGN 268

Qy     134 QRPLLDSVAQQQGSSLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQLKLTDGSAH-RFEV 192
           :|                :|   | :    : :|: |       ||  ||   : |:
Db     269 KRK-----------------KWYRPVTVQQLLEIKSIHPDA----KLIGGSTETQIEI 305

Qy     193 PIAKFQELRYSVALVLKEMAELEK 216
               ||:::||  :: | ::||| :
Db     306 ---KFKQMRYGASVYLGDLAELRQ 326


RESULT   6
COBK_METTH
ID   COBK_METTH     STANDARD;      PRT;   302 AA.
AC   O27083;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
```

```
DE   PROBABLE PRECORRIN-6X REDUCTASE (EC 1.3.1.54).
GN   COBK OR MTH1002.
OS   Methanobacterium thermoautotrophicum.
OC   Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC   Methanobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DELTA H;
RX   MEDLINE; 98037514.
RA   Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA   Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA   Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA   Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA   Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA   McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA   Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;
RT   "Complete genome sequence of Methanobacterium thermoautotrophicum
RT   deltaH: functional analysis and comparative genomics.";
RL   J. Bacteriol. 179:7135-7155(1997).
CC   -!- FUNCTION: CATALYZES THE REDUCTION OF THE MACROCYCLE OF PRECORRIN-
CC       6X INTO PRECORRIN-6Y (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: PRECORRIN-6Y + NADP(+) = PRECORRIN-6X + NADPH.
CC   -!- PATHWAY: COBALAMIN BIOSYNTHESIS (BY SIMILARITY).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE000873; AAB85498.1; -.
KW   Oxidoreductase; NADP; Cobalamin biosynthesis.
SQ   SEQUENCE   302 AA;  32673 MW;  F0BCDBC37DEDF3C5 CRC64;


  Query Match             7.5%;  Score 83;  DB 1;  Length 302;
  Best Local Similarity   24.5%;  Pred. No. 7.6;
  Matches   56;  Conservative   30;  Mismatches   85;  Indels   58;  Gaps   10;

Qy      22 VSFLGSQPSPEVTAVAQLLKDLDRSTFRKLLKLVVGALHGKDCREAVEQLGASANLSEER 81
           :| |   |  |||| |      | :      : | |||     || :  | |
Db      54 ISRLSGMPWVEVTATATTEHGSDLAEKSGASRTVTGALDSDGLRELMADLDA-------- 106

Qy      82 LAVLLAGTHTLLQQ----ALR----------------LPPASLKPDAFQEELQELGIPQ 120
             :|:  ||    |    |||                :|   ::  :|:|
Db     106 -CILIDATHPFAAQATENALRACRETGTIYVRFERPEVIPDGVIRVGSFRE--------- 156

Qy     121 DLIGDLASLAFGSQRPLLD-SVAQQQGSSLPHVSYFRWRVDVAISTSAQSRSLQ----PS 175
             |::||   |    ::  :     |  |  :   |  | |  |||:  : ||    ||
Db     156 --AGEVASSLIGDGEVVMHLAGVSTLGDVLRSLEPERVAVRVLPSTSSIEKCLQLGVPPS 213

Qy     176 VLMQLKLTDGSAHRF--EVPIAKFQELRYSVALVLKEMAE---LEKKCE 219
           :: ::       ||  |: :|  :| | : |:: ||  |    | :| |
Db     214 HIIAMQ------GRFSAEMNLALLREYR-AGAVITKESGETGGLPEKVE 255


RESULT   7
MIS_HUMAN
ID   MIS_HUMAN      STANDARD;      PRT;   560 AA.
AC   P03971; O75246;
DT   23-OCT-1986 (Rel. 02, Created)
DT   23-OCT-1986 (Rel. 02, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   MUELLERIAN INHIBITING FACTOR PRECURSOR (MIS) (ANTI-MUELLERIAN HORMONE)
DE   (AMH) (MULLERIAN INHIBITING SUBSTANCE).
GN   AMH OR MIF.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 86218082.
RA   Cate R.L., Mattaliano R.J., Hession C., Tizard R., Farber N.M.,
RA   Cheung A., Ninfa E.G., Frey A.Z., Gash D.J., Chow E.P., Fisher R.A.,
RA   Bertonis J.M., Torres G., Wallner B.P., Ramachandran K.L.,
RA   Ragin R.C., Manganaro T.F., McLaughlin D.T., Donahoe P.K.;
RT   "Isolation of the bovine and human genes for Mullerian inhibiting
RT   substance and expression of the human gene in animal cells.";
RL   Cell 45:685-698(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Lamerdin J.E., McCready P.M., Skowronski E., Adamson A.W.,
RA   Burkhart-Schultz K.J., Gordon L., Kyle A., Ramirez M., Stilwagen S.,
RA   Phan H., Velasco N., Do L., Regala W., Terry A., Garnes J.,
RA   Danganan L., Poundstone P., Christensen M., Georgescu A., Avila J.,
RA   Liu S., Attix C., Andreise T., Trankheim M., Amico-Keller G.,
RA   Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
RA   Kronmiller B., Arellano A., Montgomery M., Ow D., Nolan M., Trong S.,
RA   Kobayashi A., Olsen A.S., Carrano A.V.;
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   VARIANT ARG-325.
RX   MEDLINE; 93131268.
RA   Carre-Eusebe D., Imbeaud S., Harbison M., New M.I., Josso N.,
RA   Picard J.Y.;
RT   "Variants of the anti-Mullerian hormone gene in a compound
RT   heterozygote with the persistent Mullerian duct syndrome and his
RT   family.";
RL   Hum. Genet. 90:389-394(1992).
RN   [4]
RP   VARIANTS PMDS G-12; P-70; V-101; W-123; C-167; C-194 AND A-477.
RX   MEDLINE; 94214429.
RA   Imbeaud S., Carre-Eusebe D., Rey R., Belville C., Josso N.,
RA   Picard J.-Y.;
RT   "Molecular genetics of the persistent mullerian duct syndrome: a
RT   study of 19 families.";
RL   Hum. Mol. Genet. 3:125-131(1994).
CC   -!- FUNCTION: THIS GLYCOPROTEIN, PRODUCED BY THE SERTOLI CELLS OF THE
CC       TESTIS, CAUSES REGRESSION OF THE MUELLERIAN DUCT. IT ALSO IS ABLE
CC       TO INHIBIT THE GROWTH OF TUMORS DERIVED FROM TISSUES OF MUELLERIAN
CC       DUCT ORIGIN.
CC   -!- SUBUNIT: HOMODIMER, DISULFIDE-LINKED.
CC   -!- DISEASE: DEFECTS IN AMH ARE THE CAUSE OF PERSISTENT MUELLERIAN
CC       DUCT SYNDROME (PMDS); A FORM OF MALE PSEUDOHERMAPHRODITISM
CC       CHARACTERIZED BY A FAILURE OF MUELLERIAN DUCT REGRESSION IN
CC       OTHERWISE NORMAL MALES.
CC   -!- MISCELLANEOUS: ALTHOUGH IT DOES NOT COMPETE WITH EGF FOR RECEPTOR
CC       BINDING SITES, MIS CAN INHIBIT THE AUTOPHOSPHORYLATION OF THE EGF
CC       RECEPTOR IN VITRO.
CC   -!- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; K03474; AAA98805.1; -.
DR   EMBL; AC005263; AAC25614.1; -.
DR   PIR; A01397; WFHUM.
DR   MIM; 600957; -.
DR   MIM; 261550; -.
DR   PRINTS; PR00438; GFCYSKNOT.
DR   PROSITE; PS00250; TGF_BETA; 1.
DR   PFAM; PF00019; TGF-beta; 1.
KW   Growth factor; Glycoprotein; Gonadal differentiation; Signal;
KW   Pseudohermaphroditism; Disease mutation; Polymorphism.
FT   SIGNAL        1     18       POTENTIAL.
FT   PROPEP       19     25       POTENTIAL.
FT   CHAIN        26    560       MUELLERIAN INHIBITING FACTOR.
FT   DISULFID    462    526       BY SIMILARITY.
FT   DISULFID    488    557       BY SIMILARITY.
```

```
FT   DISULFID    492    559       BY SIMILARITY.
FT   DISULFID    525    525       INTERCHAIN (BY SIMILARITY).
FT   CARBOHYD     64     64       POTENTIAL.
FT   CARBOHYD    329    329       POTENTIAL.
FT   VARIANT      12     12       V -> G (IN PMDS).
FT                                /FTId=VAR_007483.
FT   VARIANT      49     49       I -> S.
FT                                /FTId=VAR_007484.
FT   VARIANT      70     70       L -> P (IN PMDS).
FT                                /FTId=VAR_007485.
FT   VARIANT     101    101       G -> V (IN PMDS).
FT                                /FTId=VAR_007486.
FT   VARIANT     123    123       R -> W (IN PMDS).
FT                                /FTId=VAR_007487.
FT   VARIANT     167    167       Y -> C (IN PMDS).
FT                                /FTId=VAR_007488.
FT   VARIANT     185    185       Q -> E.
FT                                /FTId=VAR_007489.
FT   VARIANT     194    194       R -> C (IN PMDS).
FT                                /FTId=VAR_007490.
FT   VARIANT     325    325       Q -> R.
FT                                /FTId=VAR_007491.
FT   VARIANT     477    477       V -> A (IN PMDS).
FT                                /FTId=VAR_007492.
FT   CONFLICT    515    515       A -> V (IN REF. 2).
SQ   SEQUENCE   560 AA;  59192 MW;  3EFC2EE4FECC364C CRC64;


  Query Match             7.4%;  Score 82;  DB 1;  Length 560;
  Best Local Similarity   23.5%;  Pred. No. 20;
  Matches   54;  Conservative   17;  Mismatches   67;  Indels   92;  Gaps    9;

Qy      29 PSPEVTAVAQLL---KDLDRSTFRKLLKLVV----GALHGKDCREAVEQLGASANLSEER 81
           | ||||     |   : | |   : | | |    || |    . ::  |  : ||  |
Db     170 PGPEVTVTRAGLPGAQSLCPSRDTRYLVLAVDRPAGAWRGSGLALTLQPRGEDSRLSTAR 229

Qy      82 LAVLLAG---------------------------------------------------- 89
           |  || |
Db     230 LQALLFGDDHRCFTRMTPALLLLPRSEPAPLPAHGQLDTVPFPPPRPSAELEESPPSADP 289

Qy      89 ---THTLLQQALRLPPA-------SLKPDAFQEELQELGIPQDLI--GDLASL--AFGSQ 134
              | | | :|||:|||       :| |||        | || |:   | |:|      :
Db     290 FLETLTRLVRALRVPPARASAPRLALDPDAL------AGFPQGLVNLSDPAALERLLDGE 343

Qy     135 RPLL-----------DSVAQQQGSSLPHVSYFRWRVDVAI-STSAQSRSL 172
           |||           |       :| |  :    ||   : : :|: |||
Db     344 EPLLLLLRPTAATTGDPAPLHDPTSAPWATALARRVAAELQAAAAELRSL 393


RESULT   8
PUR4_DROME
ID   PUR4_DROME     STANDARD;      PRT;  1354 AA.
AC   P35421;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   PHOSPHORIBOSYLFORMYLGLYCINAMIDINE SYNTHASE (EC 6.3.5.3) (FGAM
DE   SYNTHASE) (FORMYLGLYCINAMIDE RIBOTIDE AMIDOTRANSFERASE) (FGARAT)
DE   (ADENOSINE-2) (FGAMS) (FORMYLGLYCINAMIDE RIBOTIDE SYNTHETASE).
GN   ADE2.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R;
RX   MEDLINE; 94095139.
RA   Tiong S.Y.K., Nash D.;
RT   "The adenosine2 gene of Drosophila melanogaster encodes a
RT   formylglycineamide ribotide amidotransferase.";
RL   Genome 36:924-934(1993).
CC   -!- CATALYTIC ACTIVITY: ATP + 5'-PHOSPHORIBOSYLFORMYLGLYCINAMIDE +
CC       L-GLUTAMINE + H(2)O = ADP + ORTHOPHOSPHATE + 5'-PHOSPHORIBOSYL-
CC       FORMYLGLYCINAMIDINE + L-GLUTAMATE.
CC   -!- PATHWAY: FOURTH STEP IN DE NOVO PURINE BIOSYNTHESIS.
CC   -!- SIMILARITY: TO E.COLI ENZYME (PURL).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U00683; AAC46468.1; -.
DR   FLYBASE; FBgn0000052; ade2.
DR   PFAM; PF00586; AIRS; 1.
KW   Purine biosynthesis; Ligase.
SQ   SEQUENCE   1354 AA;  148195 MW;  30AF726DFF0353A0 CRC64;


  Query Match             7.3%;  Score 81.5;  DB 1;  Length 1354;
  Best Local Similarity   23.5%;  Pred. No. 65;
  Matches   61;  Conservative   34;  Mismatches   84;  Indels   81;  Gaps   13;

Qy      17 SHSGRVSFLGSQP-----SPEVTA---VAQLLKDLDRSTFRKLLKLV---------- 56
           |||| : :|:||       |   |   ||: | :|    | |: :|
Db     725 SHSGIATSIGTQPLKGLLDPAAMARMCVAEALSNL---VFVKISELADVKCSGNWMWAAK 781

Qy      56 ---VGALHGKDCREAVEQLGASANLSEERLAVLLAGTHTLLQQALRLPPASLK------- 106
              ||     |:|  : |        | | : : |    |  | ::   ::|
Db     782 LPGEGARMFDACKELCQIL--------EELHIAIDGGKDSLSMAAKVGGETIKSPGTLVI 833

Qy     106 ------PDAFQEELQELGIPQDLIGDLASLAFGSQRPLL-----------DSVAQ---Q 144
                 ||        | :  || |  |    ||:  ||             ::||   |
Db     834 STYAPCPDV------RLKVTPDLKGPGA----GSKTSLLWINLENSARLGGSALAQAYAQ 883

Qy     145 QGSSLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQL--KLTDGS--AHRFEVPIAKFQEL 200
           ||   |:::    | ||      | ::||    |:|    ::||       |: |     |
Db     884 QGKDTPNLT----RSDVLGKAFAVTQSLLGDGLIQAGHDVSDGGLLVCVLEMAIGGLSGL 939

Qy     201 RYSVALVLKEMAELEKKCER 220
           |  ::  | ::   :|  |:
Db     940 RVDLSEPLAKLKNFDKSVEK 959


RESULT   9
ACCA_HAEIN
ID   ACCA_HAEIN     STANDARD;      PRT;   315 AA.
AC   P43872;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   ACETYL-COENZYME A CARBOXYLASE CARBOXYL TRANSFERASE SUBUNIT ALPHA
DE   (EC 6.4.1.2).
GN   ACCA OR HI0406.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RD / KW20;
RX   MEDLINE; 95350630.
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus
RT   influenzae Rd.";
```

```
RL   Science 269:496-512(1995).
CC   -!- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE ACETYL COENZYME A
CC       CARBOXYLASE COMPLEX; FIRST, BIOTIN CARBOXYLASE CATALYZES THE
CC       CARBOXYLATION OF THE CARRIER PROTEIN AND THEN THE TRANSCARBOXYLASE
CC       TRANSFERS THE CARBOXYL GROUP TO FORM MALONYL-COA (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: CARBOXYBIOTIN CARBOXYL CARRIER PROTEIN +
CC       ACETYL-COA = BIOTIN CARBOXYL CARRIER PROTEIN + MALONYL-COA.
CC   -!- PATHWAY: FIRST STEP IN LONG-CHAIN FATTY ACID SYNTHESIS.
CC   -!- SUBUNIT: ACETYL-COA CARBOXYLASE IS AN HETEROHEXAMER OF BIOTIN
CC       CARBOXYL CARRIER PROTEIN, BIOTIN CARBOXYLASE AND THE TWO SUBUNITS
CC       OF CARBOXYL TRANSFERASE IN A 2:2 COMPLEX (BY SIMILARITY).
CC   -!- SIMILARITY: TO THE C-TERMINUS OF MAMMALIAN PROPIONYL-COA
CC       CARBOXYLASE BETA CHAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32724; AAC22065.1; -.
DR   TIGR; HI0406; -.
DR   PRINTS; PR01069; ACCCTRFRASEA.
KW   Fatty acid biosynthesis; Ligase.
SQ   SEQUENCE   315 AA;  35126 MW;  E3B0B498A2686B23 CRC64;


  Query Match             7.2%;  Score 80.5;  DB 1;  Length 315;
  Best Local Similarity   25.8%;  Pred. No. 13;
  Matches   61;  Conservative   28;  Mismatches   74;  Indels   73;  Gaps   12;

Qy      12 HHPADSHSGRVSFLGSQPSPEVTAVAQLLKDLDRSTFRKLLKLVVGALHGK-------DC 64
           |    |   :|      |:||               :|| |:|:  |   |       |
Db     114 HQKGRSVKEKVQRNFGMPAPE-------------GYRKALRLMEMAERFKLPIITFIDT 159

Qy      65 REAVEQLGASANLSEERLAVLLAGTHTLLQQALRLPPASLKPDAFQEELQELGIPQ--DL 122
             |   :||     |||      |    : : ||           |: :| :|    :
Db     160 PGAYPGIGA-----EER------GQAEAIARNLR------------EMAQLTVPVICTV 195

Qy     123 IGDLAS---LAFGSQRPLLDSVAQQQGSSLPHVS-----YFRWRVDVAISTSAQSRSLQP 174
           ||:  |   || |    : | |   | |:   :|          |:     ||:|:   |
Db     196 IGEGGSGGALAIG----VGDKVNMLQYSTYSVISPEGCASILWKSAEKASTAAEVMGLTA 251

Qy     175 SVLMQLKLTD-------GSAHRFEVPIAKFQELRYSVALVLKE-MAELEKKCERKL 222
           | | :| | |       | |||  : ||:  :||      ||| :|||::  : :|
Db     252 SRLKELNLIDSIVQEPLGGAHRSYLEIAENLKLR------LKEDLAELDELSKEEL 301


RESULT  10
POL_MLVRD
ID   POL_MLVRD      STANDARD;      PRT;  1196 AA.
AC   P11227;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1989 (Rel. 11, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   POL POLYPROTEIN [CONTAINS: PROTEASE (EC 3.4.23.-); REVERSE
DE   TRANSCRIPTASE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)].
GN   POL.
OS   Radiation murine leukemia virus.
OC   Viruses; Retroid viruses; Retroviridae; Mammalian type C retroviruses.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 87207680.
RA   Merregaert J., Janowski M., Reddy E.P.;
RT   "Nucleotide sequence of a radiation leukemia virus genome.";
RL   Virology 158:88-102(1987).
CC   -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC   -!- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-POL
CC       POLYPROTEIN.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A2; ALSO KNOWN AS THE
CC       RETROPEPSIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; K03363; -; NOT_ANNOTATED_CDS.
DR   PIR; B26183; GNMVRV.
DR   HSSP; P03355; 1MML.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PFAM; PF00075; rnaseH; 1.
DR   PFAM; PF00077; rvp; 1.
DR   PFAM; PF00078; rvt; 1.
DR   PFAM; PF00552; integrase; 1.
DR   PFAM; PF00665; rve; 1.
KW   Hydrolase; Transferase; RNA-directed DNA polymerase;
KW   Aspartyl protease; Endonuclease; Polyprotein.
FT   CHAIN         1    103       PROTEASE.
FT   ACT_SITE     27     27       BY SIMILARITY.
SQ   SEQUENCE   1196 AA;  133322 MW;  9549253FCEEB5306 CRC64;


  Query Match             7.2%;  Score 80.5;  DB 1;  Length 1196;
  Best Local Similarity   24.2%;  Pred. No. 67;
  Matches   54;  Conservative   30;  Mismatches   84;  Indels   55;  Gaps   10;

Qy      33 VTAVAQLLKDLDRSTFRKLLKLVVGALHGKDCREAVEQLGASANLSEERL----AVLLAG 88
           | |:| | ||  : |  :   ||: | |     ||: :       ||  |:    |:||
Db     533 VAAIAVLTKDAGKLTMGQ--PLVILAPHAV---EALVKQPPDRWLSNARMTHYQAMLLDT 587

Qy      89 THTLLQQALRLPPASLKPDAFQEELQELGIPQDLIGDLASLAFGSQRPLLDSV------- 142
                 : | ||:| |       | | | | | :  ||    |::  | |
Db     588 DRVQFGPVVALNPATLLP------LPEEGAPHDCLEILAE-THGTEPDLTDQPIPDADHT 640

Qy     142 ------------AQQQGSSLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQLKLTDGSAHR 189
                       :: |:::   :   |   :   ||||   |  ::   ||: :|   |
Db     641 WYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSAQRAEL-IALTQALKMAEGK--R 697

Qy     190 FEVPIAKFQELRYSVA------------LVLKEMAELEKKCE 219
             |    : : ||: |             |: |  |:: | |
Db     698 LNV----YTDSRYAFATAHIHGEIYKRRGLLTSEGREIKNKSE 736


RESULT  11
DYNA_RAT
ID   DYNA_RAT       STANDARD;      PRT;  1280 AA.
AC   P28023;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   DYNACTIN, 150 KD ISOFORM (150 KD DYNEIN-ASSOCIATED POLYPEPTIDE)
DE   (DP-150) (DAP-150) (P150-GLUED).
GN   DCTN1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY;
RX   MEDLINE; 91260877.
RA   Holzbaur E.L.F., Hammarback J.A., Paschal B.M., Kravit N.G.,
RA   Pfister K.K., Vallee R.B.;
RT   "Homology of a 150K cytoplasmic dynein-associated polypeptide with
RT   the Drosophila gene Glued.";
RL   Nature 351:579-580(1991).
RN   [2]
RP   REVISIONS.
RA   Holzbaur E.L.F., Hammarback J.A., Paschal B.M., Kravit N.G.,
RA   Pfister K.K., Vallee R.B.;
RL   Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
```

```
CC   -!- FUNCTION: REQUIRED FOR THE CYTOPLASMIC DYNEIN-DRIVEN RETROGRADE
CC       MOVEMENT OF VESICLES AND ORGANELLES ALONG MICROTUBULES. DYNEIN-
CC       DYNACTIN INTERACTION IS A KEY COMPONENT OF THE MECHANISM OF AXONAL
CC       TRANSPORT OF VESICLES AND ORGANELLES.
CC   -!- SUBUNIT: LARGE MACROMOLECULAR COMPLEX OF AT LEAST 10 COMPONENTS.
CC       P150(GLUED) BINDS DIRECTLY TO MICROTUBULES AND TO CYTOPLASMIC
CC       DYNEIN.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- SIMILARITY: CONTAINS 1 CAP-GLY DOMAIN.
CC   -!- SIMILARITY: STRONG, TO OTHER SPECIES DYNACTIN 150 KD SUBUNIT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62160; CAA44091.1; -.
DR   PIR; S16129; S16129.
DR   PROSITE; PS00845; CAP_GLY; 1.
DR   PFAM; PF01302; CAP_GLY; 1.
KW   Motor protein; Microtubules; Dynein; Coiled coil; Cytoskeleton.
FT   DOMAIN       48     90       CAP-GLY.
FT   DOMAIN      157    184       SER-RICH.
FT   DOMAIN      214    513       COILED COIL (POTENTIAL).
FT   DOMAIN      942   1048       COILED COIL (POTENTIAL).
FT   DOMAIN     1184   1213       COILED COIL (POTENTIAL).
SQ   SEQUENCE   1280 AA;  141929 MW;  C9348CF129F4FF5C CRC64;


  Query Match             7.2%;  Score 80.5;  DB 1;  Length 1280;
  Best Local Similarity   23.7%;  Pred. No. 73;
  Matches   58;  Conservative   34;  Mismatches   90;  Indels   63;  Gaps   11;

Qy      20 GRV-SFLGSQPSPEVTAVAQLLKDLDR--STFRKLLKLVVGALHGKDCREAVEQLGASAN 76
           ||: :|| |   | | :| ||:||:   | |:  | :   : | |       |   :
Db     764 GRLRAFL--QGGQEATDIALLLRDLETSCSDIRQFCKKIRRRMPGTDAPGIPAALAFGSQ 821

Qy      77 LSE----------------ERLAVLLAGTHTLLQQALRLPPASLKPDAFQEELQELGIPQ 120
           :|:                : :|   |     | :   || |:|:  ||:   |  | |
Db     822 VSDTLLDCRKHLTWVVAVLQEVAAAAAQLIAPLAENEGLPVAALEELAFKASEQIYGSPS 881

Qy     121 D------------LIGDLASLAFGSQRPLLDSVAQQQGSSLPHVSYFRWRVDVAISTSAQ 168
                       ||  :  ||   |    |  |::  |  | |    |
Db     882 SSPYECLRQSCSILISTMNKLATAMQEGEYD--AERPPSKPPPVE--PW---------- 927

Qy     169 SRSLQPSVLMQLKLTD--GSAHRFEVPIAKFQELRYSVALVLKEMAE-------LEKKCE 219
             |: | : ::||  |   : |      :||: |: :  :|::|      ||||  :
Db     927 -----PAAL-RAEITDAEGLGLKLEDRETVIKELKKSLKIKGEELSEANVRLSLLEKKLD 980

Qy     220 RKLQD 224
             :|
Db     981 SAAKD 985


RESULT  12
P152_YEAST
ID   P152_YEAST     STANDARD;      PRT;  1337 AA.
AC   P39685;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   NUCLEAR ENVELOPE PORE MEMBRANE PROTEIN POM152 (P150).
GN   POM152 OR YMR129W OR YM9553.05.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=W303;
RX   MEDLINE; 94186543.
RA   Wozniak R.W., Blobel G., Rout M.P.;
RT   "POM152 is an integral protein of the pore membrane domain of the
RT   yeast nuclear envelope.";
RL   J. Cell Biol. 125:31-42(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Badcock K., Churcher C., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL   Submitted (MAR-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: COMPONENT OF THE NUCLEAR PORE COMPLEX WHICH MAY PROVIDE
CC       THE NECESSARY ASYMMETRY REQUIRED FOR ANCHORING STRUCTURES SUCH AS
CC       CYTOPLASMIC FILAMENTS AND THE NUCLEOPLASMIC CAGE.
CC   -!- SUBUNIT: INTERACTS WITH NUP170 AND/OR NUP188.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. NUCLEAR PORE
CC       MEMBRANE.
CC   -!- PTM: THE N-TERMINUS IS BLOCKED.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z31592; CAA83469.1; -.
DR   EMBL; Z48622; CAA88554.1; -.
DR   PIR; A53824; A53824.
DR   SGD; L0001464; POM152.
KW   Nuclear protein; Transmembrane; Repeat; Glycoprotein.
FT   DOMAIN        1    175       PORE SIDE (POTENTIAL).
FT   TRANSMEM    176    195       POTENTIAL.
FT   DOMAIN      196   1337       CISTERNAL SIDE (POTENTIAL).
FT   CARBOHYD    280    280
FT   DOMAIN      390   1276       8 X 24 AA APPROXIMATE REPEATS.
FT   REPEAT      390    413       1.
FT   REPEAT      626    650       2.
FT   REPEAT      732    755       3.
FT   REPEAT      836    859       4.
FT   REPEAT      943    966       5.
FT   REPEAT     1058   1077       6.
FT   REPEAT     1157   1178       7.
FT   REPEAT     1253   1276       8.
SQ   SEQUENCE   1337 AA;  151651 MW;  A024A42069193898 CRC64;


  Query Match             7.2%;  Score 80.5;  DB 1;  Length 1337;
  Best Local Similarity   23.7%;  Pred. No. 77;
  Matches   42;  Conservative   24;  Mismatches   48;  Indels   63;  Gaps    8;

Qy      71 LGASANLSEERLAV---LLAGTHTLLQQALRLPPASLKPDAFQEELQELGIPQDLI-GDL 126
           :| | |:| |    |    || :  :  |      :||:|: |:  ||     |    |:|
Db     393 IGDSDNVSFEIQGVPPMKLAYSKIVNGQTFSYVDSSLQPEYFESPLQSSKSKQSFTQGEL 452

Qy     127 ASLAFGSQRPL---LDSVAQQQGSSLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQLKLT 183
             | :|  :|:   |||   | |        | :::|                    |:|
Db     453 NDLKWGRNQPVNINLDSSITQDGK-------FAYKID-------------------KIT 485

Qy     184 DG------------------SAHRF---EVPIAKFQE-------LRYSVALVLKEM 211
           ||                    :: |   ||| |  :|        : |:|:| :|:
Db     486 DGLGNVVDFTSLPEELKKRYDLSYNFNVHEVPRAALEERFDPKSPTKRSIAIVFEEI 542


RESULT  13
TALA_DICDI
ID   TALA_DICDI     STANDARD;      PRT;  2491 AA.
AC   P54633;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   FILOPODIN (TALIN HOMOLOG).
GN   TALA.
OS   Dictyostelium discoideum (Slime mold).
```

```
OC   Eukaryota; Dictyosteliida; Dictyostelium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX2;
RX   MEDLINE; 95213284.
RA   Kreitmeier M., Gerisch G., Heizer C., Mueller-Taubenberger A.;
RT   "A talin homologue of Dictyostelium rapidly assembles at the leading
RT   edge of cells in response to chemoattractant.";
RL   J. Cell Biol. 129:179-188(1995).
CC   -!- FUNCTION: ACTIN-BINDING PROTEIN THAT MAY BE INVOLVED IN THE
CC       CONTROL OF CELL MOTILITY AND CHEMOTAXIS.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC; RAPIDLY ASSEMBLES AT THE
CC       LEADING EDGE OF CELLS IN RESPONSE TO CHEMOATTRACTANT.
CC   -!- SIMILARITY: CONTAINS A DOMAIN FOUND IN BAND 4.1, EZRIN, MOESIN,
CC       RADIXIN, AND TALIN.
CC   -!- SIMILARITY: IN THE C-TERMINUS TO YEAST SLA2 AND C.ELEGANS ZK370.3.
CC       PARTIAL, TO YEAST ROD1.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U14576; AAC46586.1; -.
DR   DICTYDB; DD01121; TALA.
DR   PROSITE; PS00660; BAND_41_1; 1.
DR   PROSITE; PS00661; BAND_41_2; 1.
DR   PROSITE; PS50057; BAND_41_3; 1.
DR   PFAM; PF00373; Band_41; 1.
DR   PFAM; PF01608; I_LWEQ; 1.
KW   Structural protein; Cytoskeleton; Actin-binding.
FT   DOMAIN      134    294       BAND 4.1-LIKE DOMAIN.
SQ   SEQUENCE   2491 AA;  268811 MW;  B24FF0224F2A419B CRC64;


  Query Match             7.2%;  Score 80.5;  DB 1;  Length 2491;
  Best Local Similarity   23.3%;  Pred. No. 1.7e+02;
  Matches   52;  Conservative   31;  Mismatches   77;  Indels   63;  Gaps    9;

Qy      11 LHHPADSHSGRVSFLGSQPSPEVTAVAQLLKDLDR-------------------STFRK 50
           | |   | :  |  : |:        ::| ||||:                    : | |
Db    1483 LAHATSSLARAVKSVTSKEPGAKAMISQSLKDLESAINNLLITSSVPASERGIGIADFNK 1542

Qy      51 LL-----------KLVVGALHGKDCREAVEQLGASANLSEERLAVLLAGTHTLLQQALRL 99
           |:           :|:: |     | :  : :  |: |||   |||:  :   |:  :::
Db    1543 LMSTCRSVSTASSQLIISA---SSCSQKPKDIELSSILSEN--AVLMTNS---LKDIIKV 1594

Qy     100 PPASLKPDAFQEELQELGIPQDLIGDLASLAFGSQRPLLDSVAQQQGSSLPHVSYFRWRV 159
             : :    | ||  |  | |  | ||:|:|       || |  :   | ||      |
Db    1595 TSSMMPGVNFCEEAIE--IAQRAISDLSSVALSVAVGSFDSSANNK-EGLSHVESQERLV 1651

Qy     160 DV-------------------AISTSAQSRS-LQPSVLMQLK 181
           ||                   ||  ||:: | : ||::   |
Db    1652 DVTKKIGTGINDLLKASRQSPEAIGISAKALSFIAPSLVNTTK 1694


RESULT  14
IUNH_CRIFA
ID   IUNH_CRIFA     STANDARD;      PRT;   314 AA.
AC   Q27546;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   INOSINE-URIDINE PREFERRING NUCLEOSIDE HYDROLASE (EC 3.2.2.1)
DE   (IU-NUCLEOSIDE HYDROLASE) (PURINE NUCLEOSIDASE).
GN   IUNH.
OS   Crithidia fasciculata.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Crithidia.
RN   [1]
RP   SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND MUTAGENESIS.
RX   MEDLINE; 96214502.
RA   Gopaul D.N., Meyer S.L., Degano M., Sacchettini J.C., Schramm V.L.;
RT   "Inosine-uridine nucleoside hydrolase from Crithidia fasciculata.
RT   Genetic characterization, crystallization, and identification of
RT   histidine 241 as a catalytic site residue.";
RL   Biochemistry 35:5963-5970(1996).
RN   [2]
RP   X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX   MEDLINE; 96214503.
RA   Degano M., Gopaul D.N., Scapin G., Schramm V.L., Sacchettini J.C.;
RT   "Three-dimensional structure of the inosine-uridine nucleoside N-
RT   ribohydrolase from Crithidia fasciculata.";
RL   Biochemistry 35:5971-5981(1996).
CC   -!- FUNCTION: CATALYZES THE HYDROLYSIS OF ALL OF THE COMMONLY
CC       OCCURRING PURINE AND PYRIMIDINE NUCLEOSIDES INTO RIBOSE AND THE
CC       ASSOCIATED BASE, BUT HAS A PREFERENCE FOR INOSINE AND URIDINE AS
CC       SUBSTRATES.
CC   -!- CATALYTIC ACTIVITY: A N-D-RIBOSYLPURINE + H(2)O = A PURINE + D-
CC       RIBOSE.
CC   -!- PATHWAY: PURINE SALVAGE.
CC   -!- SUBUNIT: HOMOTETRAMER.
CC   -!- MASS SPECTROMETRY: MW=34194; MW_ERR=4; METHOD=ELECTROSPRAY.
CC   -!- SIMILARITY: BELONGS TO THE IUNH FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U43371; AAC47119.1; -.
DR   PDB; 1MAS; 17-AUG-96.
DR   PDB; 2MAS; 12-AUG-97.
DR   PROSITE; PS01247; IUNH; 1.
DR   PFAM; PF01156; IU_nuc_hydro; 1.
KW   Hydrolase; 3D-structure.
FT   INIT_MET      0      0
FT   ACT_SITE    240    240
FT   MUTAGEN     240    240       H->A: LOSS OF ACTIVITY.
SQ   SEQUENCE   314 AA;  34194 MW;  21B51B3878A73FA1 CRC64;


  Query Match             7.2%;  Score 80;  DB 1;  Length 314;
  Best Local Similarity   20.5%;  Pred. No. 14;
  Matches   46;  Conservative   30;  Mismatches   92;  Indels   56;  Gaps    6;

Qy      25 LGSQPSPEVTAVAQLLKDL------------DRSTFRKLLKLVVGALHGKDCREAVEQLG 72
           :|:|   :||  |||: |:             |:   ||:  :  | :||:    |
Db      36 VGNQTLAKVTRNAQLVADIAGITGVPIAAGCDKPLVRKI--MTAGHIHGESGMGTVAYPA 93

Qy      73 ASANLSEERLAVLL----------------AGTHTLLQQALRLPPASLKPDAFQEELQE 115
             |  :|| || |                  |  | :  | || |  :
Db      94 EFKNKVDERHAVNLIIDLVMSHEPKTITLVPTGGLTNIAMAARLEPRIV---------- 143

Qy     116 LGIPQDLIGDLASLAFGSQRPLLDSVAQQQGSSLPHVSYF----RWRVDVAISTSAQSRS 171
             | : ::  :  |       |||:      |  ::      |:| :
Db     143 -----DRVKEVVLMGGGYHEGNATSVAEFNIIIDPEAAHIVFNESWQVTMVGLDLTHQAL 197

Qy     172 LQPSVLMQLKLTDGSAHRFEVPIAK-----FQELRYSVALVLKE 210
             | :| ::|  | :  || : |       :|  ||  |  : :
Db     198 ATPPILQRVKEVDTNPARFMLEIMDYYTKIYQSNRYMAAAAVHD 241


RESULT  15
RB6K_MOUSE
ID   RB6K_MOUSE     STANDARD;      PRT;   887 AA.
AC   P97329;
DT   15-FEB-2000 (Rel. 39, Created)
DT   15-FEB-2000 (Rel. 39, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   RABKINESIN-6 (RAB6-INTERACTING KINESIN-LIKE PROTEIN) (KINESIN-LIKE
```

```
DE   PROTEIN 174).
GN   RAB6KIFL.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=TESTIS;
RX   MEDLINE; 98101856.
RA   Echard A., Jollivet F., Martinez O., Lacapere J.-J., Rousselet A.,
RA   Janoueix-Lerosey I., Goud B.;
RT   "Interaction of a Golgi-associated kinesin-like protein with Rab6.";
RL   Science 279:580-585(1998).
CC   -!- FUNCTION: INTERACTS WITH GUANOSINE TRIPHOSPHATE (GTP)-BOUND FORMS
CC       OF RAB6. MAY ACT AS A MOTOR REQUIRED FOR THE RETROGRADE RAB-6
CC       REGULATED TRANSPORT OF GOLGI MEMBRANES AND ASSOCIATED VESICLES
CC       ALONG MICROTUBLES. HAS A MICROTUBULE PLUS END-DIRECTED MOTILITY.
CC   -!- SUBCELLULAR LOCATION: GOLGI.
CC   -!- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED, WITH HIGHEST LEVELS IN
CC       SPLEEN AND TESTIS.
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y09632; CAA70845.1; -.
DR   MGD; MGI:1201682; RAB6KIFL.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; FALSE_NEG.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
DR   PFAM; PF00225; kinesin; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil; Golgi stack;
KW   Protein transport.
FT   DOMAIN       69    494       MECHANOCHEMICAL (MOTOR).
FT   NP_BIND     159    166       ATP (POTENTIAL).
FT   DOMAIN      559    804       COILED COIL (POTENTIAL).
FT   DOMAIN      805    887       GLOBULAR (POTENTIAL).
SQ   SEQUENCE   887 AA;  99876 MW;  7143CCB261A4EA15 CRC64;


  Query Match             7.2%;  Score 80;  DB 1;  Length 887;
  Best Local Similarity   22.7%;  Pred. No. 51;
  Matches   58;  Conservative   35;  Mismatches  105;  Indels   58;  Gaps   10;

Qy       9 PYLHHPADSHSGRV----------SFLGSQP----------SPEVTAVAQLLKDLDRST 47
           | ||     || :|           | |   |            |:  | : :| |
Db     517 PSLHSFIKKHSPQVGPGLEKEDKADSDLEDSPEDEADVSVYGKEELLQVVEAMKALLLKE 576

Qy      48 FRKLLKLVVGALHGKDCREAVEQLGASANLSEERLAVLLAGTHTLLQQALRLPPASLKPD 107
            :: |:| :   |  : | | |||:        |||        | :: |::   ||
Db     577 RQEKLQLEI-QLREEICNEMVEQMQQREQWCSERLDNQKELMEELYEEKLKILKESL-TT 634

Qy     108 AFQEELQELGIPQDLIGDLASLAFGSQRPLLDSVAQQQGSSLPHVSYFRWRVDVAISTSA 167
           :||::||     : | :| :|    |       ||| |     :|  |    :| | |
Db     635 FYQEQIQE---RDEKIEELETLL---QEAKQQPAAQQSGG----LSLLRRSQRLAASAST 684

Qy     168 QSRSLQPSVLMQLKL----TDGSAHRFEVPI--------------AKFQELRYSVALVLK 209
           |      : | | |     |    |:::  :               | :| : :: |:
Db     685 QQFQEVKAELEQCKTELSSTTAELHKYQQVLKPPPPAKPFTIDVDKKLEEGQKNIRLLRT 744

Qy     210 EMAEL-------EKKC 218
           |: :|       |: |
Db     745 ELQKLGQSLQSAERAC 760


Search completed: May 15, 2000, 05:34:20
Job time: 12490 sec
```

OM protein - protein search, using sw model

Run on: May 15, 2000, 02:05:54 ; Search time 55.93 Seconds
(without alignments)
277.684 Million cell updates/sec

Title: US-09-223-796-2
Perfect score: 1112
Sequence: 1 MSALGAAAPYLHHPADSHSG.....ALVLKEMAELEKKCERKLQD 224

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : SPTREMBL_12:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	124.5	11.2	202	4	Q9Y6G5	Q9y6g5 homo sapien
2	104.5	9.4	676	5	Q27512	Q27512 caenorhabdi
3	96	8.6	3027	4	Q9Y485	Q9y485 homo sapien
4	93	8.4	486	2	Q9Z6N7	Q9z6n7 chlamydia p
5	91	8.2	584	2	Q59152	Q59152 agrobacteri
6	89.5	8.0	433	5	Q93317	Q93317 caenorhabdi
7	89	8.0	3187	11	Q63714	Q63714 rattus norv
8	88	7.9	678	5	O61493	O61493 drosophila
9	85.5	7.7	336	2	Q54057	Q54057 salmonella
10	84.5	7.6	674	5	Q93343	Q93343 caenorhabdi
11	84.5	7.6	1558	5	O96275	O96275 plasmodium
12	84	7.6	303	10	P93792	P93792 triticum ae
13	84	7.6	649	4	O43642	O43642 homo sapien
14	84	7.6	790	3	Q07732	Q07732 saccharomyc
15	84	7.6	4151	5	O96936	O96936 drosophila
16	83.5	7.5	336	5	Q9XUK7	Q9xuk7 caenorhabdi
17	83.5	7.5	1583	5	O76222	O76222 trypanosoma
18	83	7.5	336	2	Q54046	Q54046 salmonella
19	83	7.5	336	2	Q54048	Q54048 salmonella
20	83	7.5	845	10	Q9ZVG0	Q9zvg0 arabidopsis
21	82.5	7.4	195	11	Q63829	Q63829 mus musculu
22	82.5	7.4	336	2	Q54061	Q54061 salmonella
23	82.5	7.4	679	5	Q24185	Q24185 drosophila
24	82.5	7.4	758	2	P94930	P94930 moraxella c
25	82.5	7.4	758	2	O66007	O66007 moraxella c
26	82.5	7.4	817	11	O35274	O35274 rattus norv
27	82.5	7.4	907	11	Q9Z1P4	Q9z1p4 mus musculu
28	81.5	7.3	734	2	Q9X112	Q9x112 thermotoga
29	81	7.3	459	5	O62068	O62068 caenorhabdi
30	81	7.3	868	12	O73505	O73505 porcine end
31	81	7.3	970	10	O22941	O22941 arabidopsis
32	81	7.3	1289	4	Q9Y2A5	Q9y2a5 homo sapien
33	81	7.3	1295	5	Q22257	Q22257 caenorhabdi
34	81	7.3	1386	4	O15028	O15028 homo sapien
35	80.5	7.2	506	8	O99015	O99015 prototheca
36	80.5	7.2	1479	11	Q9Z277	Q9z277 mus musculu
37	80	7.2	818	4	O75160	O75160 homo sapien
38	80	7.2	998	4	Q12869	Q12869 homo sapien
39	80	7.2	1324	4	Q15312	Q15312 homo sapien
40	80	7.2	1638	4	Q9Y6J4	Q9y6j4 homo sapien
41	80	7.2	1712	4	Q9Y6J5	Q9y6j5 homo sapien
42	80	7.2	1961	11	Q62812	Q62812 rattus norv
43	80	7.2	2319	4	Q13023	Q13023 homo sapien
44	79.5	7.1	397	12	Q83497	Q83497 murine leuk
45	79.5	7.1	672	12	Q83379	Q83379 rat leukemi

ALIGNMENTS

RESULT 1
Q9Y6G5
ID Q9Y6G5 PRELIMINARY; PRT; 202 AA.
AC Q9Y6G5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE PTD002.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PITUITARY TUMOR;
RA FU G., HUANG Q., SONG H., PENG J., ZHANG Q., MAO M., DAI M., MAO Y.,
RA ZHOU J., CHEN Z., CHEN J.;
RT "Human PTD002 mRNA, complete cds.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF078857; AAD44489.1; -.
SQ SEQUENCE 202 AA; 22966 MW; 51317189 CRC32;

Query Match 11.2%; Score 124.5; DB 4; Length 202;
Best Local Similarity 22.2%; Pred. No. 0.0037;
Matches 45; Conservative 45; Mismatches 78; Indels 35; Gaps 6;

```
Qy      30 SPEVTAVAQLLKDLDRSTFRKLLKLVVGALHGKD----CREAVEQLGASANLSEERLAVL 85
           || :     |: :|   | :|| ::  || |       |  |:| |: :| :: | ::
Db      12 SPSMKKAVSLINAIDTGRFPRLLTRILQKLHLKAESSFSEEEEEKLQAAFSLEKQDLHLV 71

Qy      86 LAGTHTLLQQALRLPPASLKPDAFQEELQELGIPQD----LIGDLASLAFGS-----QRP 136
           |     :|:||:     ::|| | |::|: : : ||     :   :|:   :      ||
Db      72 LETISFILEQAVY---HNVKPAALQQQLENIHLRQDKAEAFVNTWSSMGQETVEKFRQRI 128

Qy     137 LLDSVAQQQGSSLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQLKLTDGSAHRFEVPIAK 196
           |    :  |          |:::: :: |||::   |  ::|| : :  :   |  : :
Db     129 LAPCKLETVG----------WQLNLQMAHSAQAKLKSPQAVLQLGVNNEDSKSLEKVLVE 178

Qy     197 FQELRYSVALVLKEMAELEKKCE 219
           |           ||: :   | |
Db     179 FSH---------KELFDFYNKLE 192
```

```
RESULT   2
Q27512
ID   Q27512      PRELIMINARY;      PRT;   676 AA.
AC   Q27512;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   NEX-2 PROTEIN.
GN   NEX-2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BERKS M.;
RL   Submitted (JAN-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
CC   -!- FUNCTION: CALCIUM/PHOSPHOLIPID-BINDING PROTEIN WHICH PROMOTES
CC       MEMBRANE FUSION AND IS INVOLVED IN EXOCYTOSIS.
CC   -!- SIMILARITY: TO OTHER PROTEINS OF THE ANNEXIN FAMILY.
DR   EMBL; Z29443; CAA82571.1; -.
DR   HSSP; P79134; 1AVC.
DR   PROSITE; PS00223; ANNEXIN; 1.
DR   PFAM; PF00191; annexin; 4.
DR   PRINTS; PR00196; ANNEXIN.
KW   Annexin; Calcium/phospholipid-binding; Repeat.
SQ   SEQUENCE   676 AA;  74969 MW;  00E36C79 CRC32;


  Query Match             9.4%;  Score 104.5;  DB 5;  Length 676;
  Best Local Similarity  25.7%;  Pred. No. 0.84;
  Matches   47;  Conservative   34;  Mismatches   53;  Indels   49;  Gaps   10;

Qy      43 LDRSTFRKLLKLVVGALHGKDCREAVEQLGASANLSEERLAVL--LAGTHTLLQQALRLP 100
           |: | |||| :||      |: |       :   ||||:|| |   | | | |
Db      50 LNNSKFRKLCELV------KESR-------SHEMLSEEQLATLYDCATTSTRL------- 90

Qy     101 PASLKPDAFQEELQELGIPQDLI--GDL--ASLAFGS------------QRPLLDSVAQ 143
                 |: | :: ::|   :::|  |:|  | | |             ::  |: : :
Db      90 -----PNDFADDQEDL---KNIIRYGELFKACHAINSTDFIQKSEDLKDEEKVALERIVE 141

Qy     144 QQ-GSSLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQLKLTDGSAHRFEVPIAKFQELRY 202
           |:   |  :    | |:: :: |  ::||:| : :|:     |  : ||  |  | : |
Db     142 QKLAESAKNEKDIAWNVNIVVANSYVAKSLRPVINIQMPTVGGDTN-FEFDIDSFAQFRQ 200

Qy     203 SVA 205
             :|
Db     201 QLA 203


RESULT   3
Q9Y485
ID   Q9Y485      PRELIMINARY;      PRT;  3027 AA.
AC   Q9Y485;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   X-LIKE 1 PROTEIN.
GN   XL1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   KRAEMER C., SCHMIDT E.R.;
RT   "Molecular analysis of the human DMXL1 gene, one of the homologues to
RT   DmX from Drosophila melanogaster.";
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ005821; CAA06718.1; -.
SQ   SEQUENCE   3027 AA;  337828 MW;  FDE10068 CRC32;


  Query Match             8.6%;  Score 96;  DB 4;  Length 3027;
  Best Local Similarity  23.4%;  Pred. No. 31;
  Matches   65;  Conservative   29;  Mismatches   74;  Indels  110;  Gaps   13;

Qy       9 PYLHHPA----DSHSGRVSFLGSQP-----------SPEVTAVAQLLKDLDRSTFRKLL 52
           |:|   |    : :|| :  |  ||            :| |      |:     |   ||
Db    1770 PFLRSMAYWILEDYSGALETLIKQPIRENDDQVLSASNPTVFNFYNYLR-----THPLLL 1824

Qy      53 KLVVG---------ALHGKDCREAVEQLGASANLSEERL------AVLLAGTHTLLQQAL 97
           :   |          :| ||         |  : |||| ||      | | ||   | : |
Db    1825 RRHFGSSDTFSTHMSLTGK------SGLAGTINLSERRLFFTTASAHLKAGCPMLALEVL 1878

Qy      98 R------------------------LPPASLKPDAFQEELQELGIPQDLIGDLASLAFG 132
                                     | : || || :::   :   | ||   | : |
Db    1879 SKMPKVIKKTRPFYRASSFLDTSKDCSPSSPLKLDAREDKSSAVDWSQSLINGFGSSSEG 1938

Qy     133 SQRPLLDSVAQQQGSSLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQ---LKL---TDGS 186
           |         :|  |:|     | |               ||||: |   |:|   :|
Db    1939 SSE-------KQSNSTLS----FDWS-------------QPSVVFQDDSLELKWDSDND 1973

Qy     187 AHRFEVPIAKFQELRYSVALVLKEMAELEKKCERKLQD 224
               :|||:            :||:  |::| ::|| |
Db    1974 EENEDVPIS-----------MKELKPLQRKTDKKLDD 1999


RESULT   4
Q9Z6N7
ID   Q9Z6N7      PRELIMINARY;      PRT;   486 AA.
AC   Q9Z6N7;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   CT863 HYPOTHETICAL PROTEIN.
GN   CPN1022.
OS   Chlamydia pneumoniae.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CWL029;
RA   KALMAN S., MITCHELL W., MARATHE R., LAMMEL C., FAN J., OLINGER L.,
RA   GRIMWOOD J., DAVIS R.W., STEPHENS R.S.;
RT   "Comparative Genomes of Chlamydia pneumoniae and C. trachomatis.";
RL   Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE001682; AAD19159.1; -.
SQ   SEQUENCE   486 AA;  54694 MW;  A5A0ABB0 CRC32;


  Query Match             8.4%;  Score 93;  DB 2;  Length 486;
  Best Local Similarity  21.9%;  Pred. No. 4.9;
  Matches   54;  Conservative   35;  Mismatches  100;  Indels   58;  Gaps    8;

Qy      27 SQPSPEVTAVAQLLKDLDRSTFRK---LLKLVVG-------ALHGKDCREAVEQLGASAN 76
           : | |      | :    | ||      :||   |       || |    :|:  |
Db       2 NMPVPSAVPSANITLKEDSSTVSTASGILKTATGEVLVSCTALEGSSSTDALISLALGQI 61

Qy      77 LSEERLAVLLAGTHTLLQQALRLPP--------------------ASLKPDAFQ-EELQ 114
           :   :  :||  |:  : | | |||                     : :|   | :
Db      62 ILATQQELLLQSTN--VHQLLFLPPEVVELEIQVVDLLVQLEHAETITSEPQETQTQSRS 119
```

```
Qy    115 ELGIPQDLIGDLASLAFGSQRPLLDSVAQQQGSSLPHVSYFR----------WRVDVAI 163
Db    120 EQTLPQQSSSKQSALSPRSLKPEISDSKQQQALQTPKDSAVRKHSEAPSPETQARASLSQ 179

Qy    164 STSAQSRSLQP-------SVLMQLKLTDGSAHRFEVPIAKFQELRYSVALVLKEMAELEK 216
Db    180 ASSSSQRSLPPQESAPERTLLEQQKASSFS------PLSQFSAEKQKEALTTSKSHELYK 233

Qy    217 KCERKLQ 223
Db    234 ERDQDRQ 240


RESULT   5
Q59152
ID   Q59152      PRELIMINARY;      PRT;   584 AA.
AC   Q59152;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   CELD PROTEIN.
GN   CELD.
OS   Agrobacterium tumefaciens.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Agrobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95164506.
RA   MATTHYSSE A.G., WHITE S., LIGHTFOOT R.;
RT   "Genes required for cellulose synthesis in Agrobacterium
RT   tumefaciens.";
RL   J. Bacteriol. 177:1069-1075(1995).
DR   EMBL; L38609; AAC41431.1; -.
SQ   SEQUENCE   584 AA;  65787 MW;  24782CA3 CRC32;

  Query Match             8.2%;  Score 91;  DB 2;  Length 584;
  Best Local Similarity   25.0%;  Pred. No. 9.2;
  Matches   53;  Conservative   27;  Mismatches   76;  Indels   56;  Gaps   10;

Qy     22 VSFLGSQPSPEVTAVAQLLKDLDRSTFRKLLKLVVG-----------ALHGKDCREAVE 69
Db    263 VLWFGDQP------LGALANIIDRQK-KSVLFYITGRDEDWKTPSPGLVLHGHCIRRAIE 315

Qy     70 Q----------------LG-----ASANLSEERLAVLLAGT------HTLLQQALRLPP 101
Db    316 QGFKTYDFLRGNEPYKYFFGPEEHKLSCTLFRTRSGDNLGGTLHPRSVRFVYEQALKLYK 375

Qy    102 ASLKPDAFQEELQELGIPQDLIG---DLASLAFGS---QRPLLDSVAQQQGSSLPHVSYF 155
Db    376 SGKKSEASIAFGQMLSAAPDHLGAQFGLANLAFDRGEFRKPRSPFSASSQQATNP--SSF 433

Qy    156 RWRVDVAI-STSAQSRSLQPSVLMQLKLTDGS 186
Db    434 GWRIGETLRCATALSRASEPSQVTNVRVPPGS 465


RESULT   6
Q93317
ID   Q93317      PRELIMINARY;      PRT;   433 AA.
AC   Q93317; Q93384;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   C32A3.3 PROTEIN.
GN   C32A3.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BURTON J.;
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   THOMAS K.;
RL   Submitted (FEB-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z81449; CAB03766.1; -.
DR   EMBL; Z48241; CAB03766.1; JOINED.
DR   EMBL; Z48241; CAA88286.1; -.
DR   EMBL; Z81449; CAA88286.1; JOINED.
SQ   SEQUENCE   433 AA;  50444 MW;  DB1524B0 CRC32;

  Query Match             8.0%;  Score 89.5;  DB 5;  Length 433;
  Best Local Similarity   23.6%;  Pred. No. 8.2;
  Matches   57;  Conservative   39;  Mismatches   95;  Indels   51;  Gaps   12;

Qy      6 AAAPYLHHPADSHSGRVSFLGSQ------PSPEVTAVAQLLKDLDRSTFRKLLKLVVGAL 59
Db     27 SAVTVLHRPRKKLSAFVSAMAARAASPQSPSKQITVVD--VYDLAASIGNDFEKLIDN-- 83

Qy     60 HGKDCREAVEQLGASANLSEERLAVLLAGTHTLLQQALRLPPA--SLKPDAFQEELQELG 117
Db     83 YGNECVRGIMPKVISA---LETLEAMAAGNDRENEEIMRLSKAVERLEQEKHQRNQQHLK 139

Qy    118 IPQDL----------IGDLASLA---FGSQRPLLDSVAQQQGSSLPHVSYFRWRVDVAIS 164
Db    140 FEEELEQVEKTYRKDIDDLQQMVKSLVNENRNLSTTV-----SSLPN------HADSPVS 188

Qy    165 TSAQSRSLQPSVLMQLKLTDGSAHRFEVPI---------AKFQELRYSVALVLKEMAELE 215
Db    189 TSMREADLK--MLLELK-EMSSQQKDEIKALQKDVDTYQCQVENLQNSIEKLIRQNEELL 245

Qy    216 KK 217
Db    246 RK 247


RESULT   7
Q63714
ID   Q63714      PRELIMINARY;      PRT;  3187 AA.
AC   Q63714;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   GIANTIN (GOLGI COMPLEX-ASSOCIATED PROTEIN OF 364 KD) (GCP364).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=KIDNEY;
RA   TOKI C., MISUMI Y., FUJIWARA T., SOHDA M., NISHIOKA M., IKEHARA Y.;
RT   "Identification and characterization of rat 364-kDa Golgi-associated
RT   protein recognized by autoantibodies from a patient with rheumatoid
RT   arthritis.";
RL   Cell Struct. Funct. 22:565-577(1997).
CC   -!- FUNCTION: MAY BE INVOLVED IN THE FORMATION AND/OR MAINTENANCE OF
CC       THE CHARACTERISTIC GOLGI STRUCTURE.
```

```
CC   -!- SUBUNIT: DISULFIDE-LINKED HOMODIMER.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. GOLGI MEMBRANE.
DR   EMBL; D25543; BAA05026.1; -.
KW   Golgi stack; Antigen; Coiled coil; Transmembrane.
FT   DOMAIN        1   3163       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   3164   3184       POTENTIAL.
FT   DOMAIN     3185   3187       LUMENAL (POTENTIAL).
FT   DOMAIN       47    116       COILED COIL (POTENTIAL).
FT   DOMAIN      126    398       COILED COIL (POTENTIAL).
FT   DOMAIN      418    479       COILED COIL (POTENTIAL).
FT   DOMAIN      505    544       COILED COIL (POTENTIAL).
FT   DOMAIN      632    891       COILED COIL (POTENTIAL).
FT   DOMAIN      924    993       COILED COIL (POTENTIAL).
FT   DOMAIN     1010   1076       COILED COIL (POTENTIAL).
FT   DOMAIN     1103   1178       COILED COIL (POTENTIAL).
FT   DOMAIN     1231   1327       COILED COIL (POTENTIAL).
FT   DOMAIN     1348   1712       COILED COIL (POTENTIAL).
FT   DOMAIN     1765   2320       COILED COIL (POTENTIAL).
FT   DOMAIN     2331   2706       COILED COIL (POTENTIAL).
FT   DOMAIN     2356   2359       POLY-GLU.
FT   DOMAIN     2723   2775       COILED COIL (POTENTIAL).
FT   DOMAIN     2800   2892       COILED COIL (POTENTIAL).
FT   DOMAIN     2953   3008       COILED COIL (POTENTIAL).
FT   DOMAIN     3069   3113       COILED COIL (POTENTIAL).
SQ   SEQUENCE   3187 AA;  364292 MW;  3C55C2B9 CRC32;
```

Query Match 8.0%; Score 89; DB 11; Length 3187;
Best Local Similarity 21.8%; Pred. No. 1.3e+02;
Matches 49; Conservative 47; Mismatches 95; Indels 34; Gaps 8;

```
Qy      11 LHHPADSHSGRVSFLGSQPSPEVTAVAQLLKDLDRSTFRKLLKLVVGALHGKDCREAVEQ 70
Db    1221 LHAAQPSHPGETATL--QATVSVAQIQDQLKEIE--VEKEELELKISSTTSELTKKSEEV 1276

Qy      71 LGASANLSEERLAVLLAGTHTLLQQALRLPPASLKPDAFQEEL-QELGIPQDLIGDLASL 129
Db    1277 LLLQEQINEQGLEI------------QNLKAASHEAKAHTEQLKQELESSQLKIADLEHL 1324

Qy     130 AFGSQRPLLDSVAQQQGSSLPHVSYF------RWRVDVAISTSAQS-----RSLQPSVLM 178
Db    1325 K--TLQPELETLQKHVGQKEEEVSYLVGQLGEKEQTLTTVQTEMEEQERLIKALHTQLEM 1382

Qy     179 QLKLTDGSAHRFEVPIAKFQELRYSVALVLKEMAELEKKCERKLQ 223
Db    1383 QAKEHEERLKQVQVEICELKKQPKE----LEEESKAKQQLQRKLQ 1423
```

```
RESULT   8
O61493
ID   O61493      PRELIMINARY;      PRT;   678 AA.
AC   O61493;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   HOOK PROTEIN.
GN   HK.
OS   Drosophila virilis (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   PHISTRY M., SUNIO A., KRAMER H.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF044926; AAC09301.1; -.
DR   FLYBASE; FBgn0024376; Dvir\hook.
SQ   SEQUENCE   678 AA;  77198 MW;  9C59D23D CRC32;
```

Query Match 7.9%; Score 88; DB 5; Length 678;
Best Local Similarity 19.2%; Pred. No. 20;
Matches 53; Conservative 54; Mismatches 75; Indels 94; Gaps 12;

```
Qy      31 PEVTAVAQLLKDLDRSTFRKLLKLVVGALHGKDCRE----------AVE 69
Db      98 PDVQAIAE---KCDLSELERLLQLVLGC--AVNCAKKQSYICEIMCLEEELQANIMRALQ 152

Qy      70 QL----------GASANLSEERLAVLLAGTHTLLQQALRLPPASLKPDA---------- 109
Db     153 ELESSTRQTTEGGVVSSLSRNSLSGMLDGNAKALEERDAMAQKCFETEKKMLLLIDEKTN 212

Qy     109 FQEELQEL-----GIPQDLIGD----LASLAFGSQR----------PLLDS------ 141
Db     213 LQQELHKLQLEFARLEHNTIGDDGVSLGPIQAGSVRYNELRRQLELVKEELLQSEGARED 272

Qy     141 ---VAQQQGSSLPHVSYFRWRVDVAISTSAQSRSLQPSVLM------QLKLTDGSAHRFE 191
Db     273 LKIKAQQQETDLLHM---QQRIDELMKSTAELTALKDEVDVLRESTDKLKVCEAQLETYK 329

Qy     192 VPIAKFQELRYSVALV-------LKEMAELEKKCER 220
Db     330 KKLEEYNDLKKHVKMLEERSADYVQQNAQFEEDAKR 365
```

```
RESULT   9
Q54057
ID   Q54057      PRELIMINARY;      PRT;   336 AA.
AC   Q54057;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   SPAN.
GN   SPAN.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S3015;
RX   MEDLINE; 97076912.
RA   BOYD E.F., WANG F.S., WHITTAM T.S., SELANDER R.K.;
RT   "Molecular genetic relationships of the salmonellae.";
RL   Appl. Environ. Microbiol. 62:804-808(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S3015;
RX   MEDLINE; 97221599.
RA   BOYD E.F., LI J., OCHMAN H., SELANDER R.K.;
RT   "Comparative genetics of the inv-spa invasion gene complex of
RT   Salmonella enterica.";
RL   J. Bacteriol. 179:1985-1991(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S3015;
RA   NELSON K., LI J.;
RL   Submitted (DEC-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U43308; AAC45006.1; -.
SQ   SEQUENCE   336 AA;  36507 MW;  D5C811AA CRC32;
```

Query Match 7.7%; Score 85.5; DB 2; Length 336;
Best Local Similarity 21.5%; Pred. No. 13;
Matches 50; Conservative 40; Mismatches 80; Indels 63; Gaps 10;

```
Qy       1 MSALGAAAPYLHHPADSHSGRVSFLGSQPSPEVTAVAQLLK----------DLDRS 46
Db      64 LAALRYGAP---AKSENHNGSVSGLHHNGKGELRIAEKLLKVTAEKSVGLVAAEAKVDKS 120

Qy      47 TF-----RKLLKLVVGALHGKDCR--EAVEQLGASAN-LSEERLAVLLAGTHTLLQQALR 98
Db     121 AVLLSSKNRPLEGVSGKKLSADLKAVESISELADNATAISDDNIKALSGDNKVIVGEGVR 180

Qy      99 LPPASLKPDAFQ----------EELQELGIPQDLIGDLASLAFGSQR---- 136
Db     181 KEGAPLARDVAPSRMAAANISKSDDKEHKKIKEASQLPLQPTTISDLSQLSGGDEKMPLA 240
```

```
Qy    136 ----------PLLDSVAQQQGSSLPHVSYFRWRVDVAISTSAQSR---SLQPS 175
                    | || | : : ||| :  : ||  | :::  |:     || ||
Db    241 AQSKPVTTIFPTLDGV-KGEDSSLTY-RFQRWGNDYSVNIQARQAGEFSLIPS 291
```

RESULT 10
Q93343
ID Q93343 PRELIMINARY; PRT; 674 AA.
AC Q93343;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE C36B1.9 PROTEIN.
GN C36B1.9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA LENNARD N.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
CC -!- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
DR EMBL; Z80215; CAB02274.1; -.
DR PROSITE; PS00518; ZINC_FINGER_C3HC4; 1.
DR PFAM; PF00097; zf-C3HC4; 1.
KW Zinc-finger.
SQ SEQUENCE 674 AA; 77728 MW; 159F7071 CRC32;

Query Match 7.6%; Score 84.5; DB 5; Length 674;
Best Local Similarity 21.2%; Pred. No. 39;
Matches 39; Conservative 30; Mismatches 60; Indels 55; Gaps 7;

```
Qy     63 DCREAVEQLGASANLSEERLAVLLAGTHTLLQQALRLPPASLKPDAFQEELQELG--IPQ 120
          :||:  ||: |:  :: : : :|   |   | : |: ||   :   : |  : ||  ||
Db    111 ECRKINEQVPATVPINFQLMQIL---TTLSLVKVLQTPPQDQEIPKY-ENFERLGTEIP- 166

Qy    121 DLIGDLASLAFGSQRPLLDSVAQQQGSSLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQL 180
            | || :|  |     | :|                 : :       |::|||:
Db    166 --IADLVNLTMGELYEHLKAVFNA---------------IRLRGKRDKTSIRPSVV--- 205

Qy    181 KLTDGSAHRFEVPIAKFQELRYSVALVLKEMA-------------------ELEKKCER 220
              ||::    |  |:  |:  ||  :                       :|:::  |
Db    205 -------HRYDSEFKKIDEVERSMNRVLHNVIKLQSGETLPRPAPTITIPYWQLDREYRR 257

Qy    221 KLQD 224
          |||
Db    258 SLQD 261
```

RESULT 11
O96275
ID O96275 PRELIMINARY; PRT; 1558 AA.
AC O96275;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE RESA-H3 ANTIGEN.
GN PFB0915W.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 99021743.
RA GARDNER M.J., TETTELIN H., CARUCCI D.J., CUMMINGS L.M., ARAVIND L.,
RA KOONIN E.V., SHALLOM S., MASON T., YU K., FUJII C., PEDERSON J.,
RA SHEN K., JING J., ASTON C., LAI Z., SCHWARTZ D.C., PERTEA M.,
RA SALZBERG S., ZHOU L., SUTTON G.G., CLAYTON R., WHITE O., SMITH H.O.,
RA FRASER C.M., ADAMS M.D., VENTER J.C., HOFFMAN S.L.;
RT "Chromosome 2 sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Science 282:1126-1132(1998).
DR EMBL; AE001424; AAC71972.1; -.
SQ SEQUENCE 1558 AA; 175658 MW; D49C2605 CRC32;

Query Match 7.6%; Score 84.5; DB 5; Length 1558;
Best Local Similarity 21.6%; Pred. No. 1.2e+02;
Matches 52; Conservative 45; Mismatches 79; Indels 65; Gaps 10;

```
Qy     27 SQPSPEVTAVAQLLKDLDRSTFRKLLKLVVGALHGKDCREAVEQLGASANLSEERLAVLL 86
          |  | | |   : :|| ::      : |||  :   |  |:||::    |: ||
Db    951 SSDSKEET---ESIKDKEKD-----VSLVVEEVQDNDMDESVEKVLELKNMEEE------ 997

Qy     87 AGTHTLLQQALRLPPASLKPDAFQEELQELG-IPQDLIGDLASL------------AFG 132
          |:: |: :   : | |   || |||  :  ||| |:  |
Db    997 -----LMKDAVEINDITSK---LIEETQELNEVEADLIKDMEKLKELEKALSEDSKEIID 1048

Qy    133 SQRPLLDSVAQQQ---GSSLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQLK-------- 182
          ::   |: | :::    ::|  |   :   :  |   :  : |:  :| ::|
Db   1049 AKDDTLEKVIEEEHDITTTLDEVVELKDVEEDKIEKVSDLKDLEEDILKEVKEIKELESE 1108

Qy    182 -----------LTDGSAHRFEVP---IAKF----QELRYSVALVLKEMAELEKKCERKLQ 223
                      ||    : |:      ||     :|::   | :|||:: || : |:||:
Db   1109 ILEDYKELKTIETDILEEKKEIEKDHFEKFEEEAEEIKDLEADILKEVSSLEVEEEKKLE 1168

Qy    224 D 224
          :
Db   1169 E 1169
```

RESULT 12
P93792
ID P93792 PRELIMINARY; PRT; 303 AA.
AC P93792;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE LOW-MOLECULAR-WEIGHT GLUTENIN STORAGE PROTEIN.
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
OC Poaceae; Triticum.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CHEYENNE;
RA ANDERSON O.D., CASSIDY B., DVORAK J.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
DR EMBL; U86028; AAB48477.1; -.
DR MENDEL; 14673; Triae;2372;14673.
DR PRINTS; PR00208; GLIADGLUTEN.
SQ SEQUENCE 303 AA; 34197 MW; C53AD656 CRC32;

Query Match 7.6%; Score 84; DB 10; Length 303;
Best Local Similarity 27.3%; Pred. No. 15;
Matches 38; Conservative 18; Mismatches 67; Indels 16; Gaps 4;

```
Qy     99 LPPA---SLKPDAFQEELQELGIPQDLIGDLASLAFGSQRPLLDS---VAQQQGSSLPHV 152
```

```
          |||    | :|   |:: | |  ||          |  |:|:|      :|||   ||
Db      38 LPPQQSFSQQPPFSQQQQQPL--PQQPSFSQQQPPFSQQQPILSQQPPFSQQQQPVLPQQ 95

Qy     153 SYFRWRVDVAISTSAQSRSL--------QPSVLMQLKLTDGSAHRFEVPIAKFQELRYSV 204
           | |  :  : :    | : |         ||||| ||        :   |:|  | |  |
Db      96 SPFSQQQQLVLPPQQQQQQLVQQQIPIVQPSVLQQLNPCKVFLQQQCSPVAMPQRLARSQ 155

Qy     205 ALVLKEMAELEKKCERKLQ 223
                  ::::| ::||
Db     156 MWQQSSCHVMQQQCCQQLQ 174


RESULT  13
O43642
ID   O43642      PRELIMINARY;      PRT;   649 AA.
AC   O43642;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   HEAT SHOCK PROTEIN 75.
GN   HSP75.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96347537.
RA   CHEN C.F., CHEN Y., DAI K., CHEN P.L., RILEY D.J., LEE W.H.;
RT   "A new member of the hsp90 family of molecular chaperones interacts
RT   with the retinoblastoma protein during mitosis and after heat shock.";
RL   Mol. Cell. Biol. 16:4691-4699(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   CHEN C.F., CHEN Y., DAI K., CHEN P.L., RILEY D.J., LEE W.H.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF043254; AAC02679.1; -.
DR   HSSP; P02829; 1AH8.
DR   PFAM; PF00183; HSP90; 2.
DR   PRINTS; PR00775; HEATSHOCK90.
KW   Heat shock.
SQ   SEQUENCE   649 AA;  74017 MW;  CCAF787C CRC32;


  Query Match             7.6%;  Score 84;  DB 4;  Length 649;
  Best Local Similarity   23.8%;  Pred. No. 40;
  Matches   46;  Conservative   39;  Mismatches   76;  Indels   32;  Gaps   10;

Qy      49 RKLLKLVVGALHGKDCREA-VEQLGASANLSEERLAVLLAGTHTLLQQALRLPPASLKPD 107
           :||| :|  :|: :  :|  : :| ::|: : |:|       | |:     ||   :
Db      43 KKLLDIVARSLYSE--KEVFIRELISNASDALEKL------RHKLVSDGQALPEMEIHLQ 94

Qy     108 AFQEE----LQELGI---PQDLIGDLASLAFGSQRPLLDSVAQQQGSSLPHVSY----FR 156
              |:    :|: ||    ::|: :| ::|    :  ||::  |  :|   :       |
Db      95 TNAEKGTITIQDTGIGMTQEELVSNLGTIARSGSKAFLDALQNQAEASSKIIGQFGVGFY 154

Qy     157 WRVDVAISTSAQSRSLQPSVLMQLKLTDGSAHRFEVPIAKFQELRYSVALVLKEMAELEK 216
               ||      |||  | |    |:|||   ||  ||:   :|    :::     |:
Db     155 SAFMVADRVEVYSRSAAPGSLGYQWLSDGSG-VFE--IAEASGVRTGTKIII----HLKS 207

Qy     217 KC-----ERKLQD 224
             |     | :::|
Db     208 DCKEFSSEARVRD 220


RESULT  14
Q07732
ID   Q07732      PRELIMINARY;      PRT;   790 AA.
AC   Q07732;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1996 (TrEMBLrel. 01, Last annotation update)
DE   CHROMOSOME IV READING FRAME ORF YDL239C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   ALT-MOERBE J., SCHNEIDER C., MORO M.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   MIPS;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z74287; CAA98819.1; -.
SQ   SEQUENCE   790 AA;  91739 MW;  6974F721 CRC32;


  Query Match             7.6%;  Score 84;  DB 3;  Length 790;
  Best Local Similarity   21.4%;  Pred. No. 52;
  Matches   57;  Conservative   41;  Mismatches   86;  Indels   82;  Gaps   12;

Qy      11 LHHPADSHSGRVSFLGSQPSPEVTAVAQL---------LKDLDRSTFRKLLKLVVGAL-- 60
           ||     || :: :|  :    :|   |           : :|    |:||   : | |
Db     517 LHFMQQQHSHKIKYLQKE--QALTKEQQFRLEKRRWHDILNLKEENFQKLKSELKGKLIL 573

Qy      60 --------------HGKDCREAVEQLGASANL------------SEERLAVLLAGTHTL 92
                         :  : :| ||:|   | :             : :::   |||  :
Db     574 SEKIQKNAEDKLNDYMNEHQEIVEKLQNQALIASRWSTQIQESENTHKKITDELAGKQS- 633

Qy      93 LQQALRLPPA--SLKPDAFQEELQELGIPQDLIGDLASLAFGSQRPLLDSVAQQQGSSLP 150
             : |:|     ||| | |||:|       : | ||          |  :   :  | | |
Db     633 --EILKLEETILSLKEDVFQEKLN----LKKLYGD----------PSTELNFETVGKSFP 676

Qy     151 HVS---YFRWRVDVAIS-TSAQSRSLQPSVLMQLKLTDGSAHRFEVPIAKFQ-------- 199
           |::   |    :|:    |  ||::|  ::|: |          ::|:  |
Db     677 HITKEKYDSLGLDILTDLTYVQSQNLIKNLLIVL----------DIPLKTFLKIVPTIVI 726

Qy     199 ELRYSVALVLKEMAELEKKCERKLQD 224
           :||  : |: |   :|  |   |   |
Db     727 QLRCELTLLTKFANDLNLKVFGKQLD 752


RESULT  15
O96936
ID   O96936      PRELIMINARY;      PRT;  4151 AA.
AC   O96936;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   GROOVIN (FRAGMENT).
GN   GRV.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 99054752.
RA   STRUMPF D., VOLK T.;
RT   "Kakapo, a novel cytoskeletal-associated protein is essential for the
RT   restricted localization of the neuregulin-like factor, vein, at the
RT   muscle-tendon junction site.";
RL   J. Cell Biol. 143:1259-1270(1998).
DR   EMBL; Y09430; CAA70581.1; -.
DR   HSSP; P07751; 1AJ3.
FT   NON_TER       1      1
SQ   SEQUENCE   4151 AA;  472264 MW;  5EEBBEC2 CRC32;


  Query Match             7.6%;  Score 84;  DB 5;  Length 4151;
  Best Local Similarity   19.1%;  Pred. No. 4.7e+02;
  Matches   56;  Conservative   51;  Mismatches  108;  Indels   78;  Gaps   10;

Qy       1 MSALGAAAPYLHHPADSHSGRVSFLG---------SQPS---PEVT-------------- 35
           :  |  | |   | |||| |  ::|           | |:    ::|
Db    2430 LGILEQALPLSEHFADSHQGLTAWLDDMEQQISRLSMPALRPDQITLQQDKNERLLQSIA 2489
```

```
Qy      35 --------------AVAQLLKDLDRSTFRKLL-----KLVVGALHGKDCREAVEQLGASA 75
                         |:  |: | | :   ::|     :     |  :: ::|:|     :
Db    2490 EHKPLLDKLNKTGEALGALVADDDGAKINEILDTDNARYAALRLELRERQQALESALQES 2549

Qy      76 NLSEERLAVLLAGTHTLLQQALRLPPASLKPDAFQEELQELGIPQDLIGDLASLAFGS-Q 134
           :   ::|  :|      : |  :| | |   |   :|::::     | : |    || : |
Db    2550 SQFSDKLEGMLRALANTVDQVNQLDPLSALPQKIREQIEDNDALMDDL-DKRQDAFSAVQ 2608

Qy     135 RPLLDSVAQQQGSSLPHVSYFRWRV--------DVAISTSAQSRSL-------QP----- 175
           |   | :|:    : | |   : ::        ||  :|  :  ||       :|
Db    2609 RAANDVIAKAGNKADPAVRDIKAKLEKLNNLWNDVQNATKKRGSSLDDILSVAEPFWKQL 2668

Qy     175 -SVLMQLKLTDGSAHRFEVPIAK----------FQELRYSVALVLKEMAELEK 216
            ||:  ||  : :    | | |:           ||:|: :     |: :: :
Db    2669 NSVMKTLKDLEETLSCQEPPAAQPQDIKKQQVALQEIRHEIDQTKPEVEQVRR 2721
```

Search completed: May 15, 2000, 05:33:25
Job time: 12451 sec

****

MPSRCH (TM)

****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 12 10:32:27 2000; MasPar time 81.38 Seconds
190.847 Million cell updates/sec
Tabular output not generated.

Title: >US-09-223-796-4
Description: (1-224) from US09223796.pep
Perfect Score: 1604
Sequence: 1 MSAVGAATPYLHHPGDSHSG.....ALVLKEMADLEKRCERRLQD 224

Scoring table: PAM 150
Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl12
1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 45.912; Variance 95.996; scale 0.478

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	125	7.8	195	11	Q63829	BUP=5'OF BMI-1 PROVIRA	1.42e-03
2	125	7.8	512	10	Q9ZW95	F5A8.3 PROTEIN.	1.42e-03
3	117	7.3	676	5	Q27512	NEX-2 PROTEIN.	1.69e-02
4	113	7.0	1576	2	O69354	RECEPTOR-LIKE HISTIDIN	5.64e-02
5	113	7.0	1576	2	O05201	RECEPTOR-LIKE HISTIDIN	5.64e-02
6	107	6.7	360	2	O33934	ERYCII.	3.27e-01
7	107	6.7	361	2	O54225	DTDP-4-KETO-6-DEOXY-HE	3.27e-01
8	107	6.7	451	2	P72863	HYPOTHETICAL 49.7 KD P	3.27e-01
9	108	6.7	463	2	O07346	PMGA,COMPLETE CDS.	2.45e-01
10	107	6.7	463	2	P73374	HYPOTHETICAL 51.5 KD P	3.27e-01
11	106	6.6	902	4	O43632	SPINDLE POLE BODY PROT	4.36e-01
12	105	6.5	678	5	O96777	CYCLIC NUCLEOTIDE AND	5.81e-01
13	104	6.5	711	11	Q9WV07	LIPOXYGENASE-3.	7.71e-01
14	103	6.4	327	2	Q9ZBT5	PUTATIVE REGULATORY PR	1.02e+00
15	102	6.4	400	2	O53544	MCE-FAMILY PROTEIN.	1.35e+00
16	103	6.4	764	4	Q9Y5S1	VANILLOID RECEPTOR-LIK	1.02e+00
17	103	6.4	764	4	Q9Y670	VANILLOID RECEPTOR-LIK	1.02e+00
18	102	6.4	1116	2	P73507	DNA POLYMERASE III SUB	1.35e+00
19	101	6.3	831	5	Q94979	TSH, FSH, LH/CG RECEPT	1.79e+00
20	100	6.2	294	2	Q44382	INNER MEMBRANE PROTEIN	2.36e+00
21	100	6.2	366	13	O93601	APOLIPOPROTEIN AIV.	2.36e+00
22	99	6.2	631	5	Q17556	SIMILARITY TO DTDP-GLU	3.10e+00
23	100	6.2	753	4	O95032	GC-RICH SEQUENCE DNA-B	2.36e+00
24	99	6.2	1010	4	O95836	MYELOID/LYMPHOID LEUKE	3.10e+00
25	100	6.2	1142	14	P89462	RIBONUCLEOTIDE REDUCTA	2.36e+00
26	98	6.1	368	2	P74288	REGULATORY COMPONENTS	4.08e+00
27	98	6.1	1198	2	Q53730	POLYKETIDE SYNTHASE (F	4.08e+00
28	96	6.0	206	11	Q62912	MYX.	6.99e+00
29	97	6.0	340	2	O84915	GALR-LACI HOMOLOG.	5.34e+00
30	97	6.0	479	2	P94426	HOMOLOGUE OF REGULATOR	5.34e+00
31	97	6.0	485	5	Q9XTV6	R12G8.2 PROTEIN.	5.34e+00
32	97	6.0	535	2	Q56676	MANNOSE-SENSITIVE HEMA	5.34e+00
33	97	6.0	649	4	O43642	HEAT SHOCK PROTEIN 75.	5.34e+00
34	96	6.0	671	2	Q9ZA89	HYPOTHETICAL 77.2 KD P	6.99e+00
35	97	6.0	709	2	P73100	COME.	5.34e+00
36	97	6.0	811	4	O75139	KIAA0644 PROTEIN.	5.34e+00
37	97	6.0	1111	4	O75155	KIAA0667 PROTEIN (FRAG	5.34e+00
38	96	6.0	2626	11	Q9Z1N3	MYOSIN-RHOGAP PROTEIN,	6.99e+00
39	95	5.9	302	9	Q37840	P2 J HOMOLOG.	9.13e+00
40	95	5.9	333	2	P71924	HYPOTHETICAL 37.1 KD P	9.13e+00
41	95	5.9	500	3	O59808	BETAINE-ALDEHYDE DEHYD	9.13e+00
42	95	5.9	563	2	P95799	RESTRICTION METHYLASE.	9.13e+00
43	95	5.9	966	13	O93338	N-METHYL-D-ASPARTATE R	9.13e+00
44	95	5.9	1557	2	O86560	HYPOTHETICAL 169.2 KD	9.13e+00
45	95	5.9	2117	2	Q9Z4Z9	CDA PEPTIDE SYNTHETASE	9.13e+00

ALIGNMENTS

RESULT 1
ID Q63829 PRELIMINARY; PRT; 195 AA.
AC Q63829;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE BUP=5'OF BMI-1 PROVIRAL INSERTION LOCUS.
GN BUP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93165034.
RA HAUPT Y., BARRI G., ADAMS J.M.;
RT "Nucleotide sequence of bup, an upstream gene in the bmi-1 proviral
RT insertion locus.";
RL Mol. Biol. Rep. 17:17-20(1992).
DR EMBL; S54914; AAB25378.1; -.
DR MGD; MGI:88218; Bup.
SQ SEQUENCE 195 AA; 22037 MW; 827A84F1 CRC32;

Query Match 7.8%; Score 125; DB 11; Length 195;
Best Local Similarity 22.2%; Pred. No. 1.42e-03;
Matches 24; Conservative 37; Mismatches 40; Indels 7; Gaps 7;

```
Db     87 LED-CKFDRERI-ELFCTEYQNNKNSLETLLGSIGRSLPHITDVSWRLEYQIKTNQLHKM 144
          |:: | : :: : :|    : : :  |:::    |  |||::|  ||::  | |: | :
Qy    113 LQELC-IPQDLVGDLASVVFGSQRPLLDSVAQQQGAWLPHVADFRWRVDVAISTSALARS 171

Db    145 YRPGYLVTLNVENNDSQSYPEINFSCNMEQLQDLVG-KLKDASKSLER 191
           :|: |: |:: :  :  : |:  : : ::|:  |:  ||: :  ||:
Qy    172 LQPSVLMQLKLSDGSAYRF-EVP-TAKFQELRYSVALVLKEMA-DLEK 216
```

RESULT 2
ID Q9ZW95 PRELIMINARY; PRT; 512 AA.
AC Q9ZW95;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE F5A8.3 PROTEIN.
GN F5A8.3.
OS Arabidopsis thaliana (Mouse-ear cress).

```
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC   Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   FEDERSPIEL N.A., PALM C.J., CONWAY A.B., KURTZ D.B., CONWAY A.R.,
RA   AU M., ARAUJO R., BUEHLER E., DEWAR K., FENG J., KIM C., LI Y.,
RA   OJI O., OSBORNE B.I., SHINN P., SUN H., TORIUMI M., VYSOTSKAIA V.S.,
RA   YU G., ECKER J., THEOLOGIS A., DAVIS R.W.;
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC004146; AAD10659.1; -.
SQ   SEQUENCE   512 AA;  58463 MW;  9C771B39 CRC32;

  Query Match             7.8%;  Score 125;  DB 10;  Length 512;
  Best Local Similarity  33.8%;  Pred. No. 1.42e-03;
  Matches    22;  Conservative   20;  Mismatches  18;  Indels   5;  Gaps   5;

Db    346 DNVRDEVRQVCGQDGVPSVEQLSSL-TSLNKVINESLRLYPPATLLPRMAFEDIKLGDLI 404
          :: || |:::     :|  |||::| :::: :: ::||| ||::| |  :| |  | :|
Qy     62 EDCRDGVQRLGVSANLPE-EQLGALLAGMHTLLQQALRL-PPTSLKPD-TFRD-QLQELC 117

Db    405 IPKGL 409
          ||  |
Qy    118 IPQDL 122


RESULT    3
ID   Q27512      PRELIMINARY;      PRT;   676 AA.
AC   Q27512;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   NEX-2 PROTEIN.
GN   NEX-2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BERKS M.;
RL   Submitted (JAN-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
CC   -!- FUNCTION: CALCIUM/PHOSPHOLIPID-BINDING PROTEIN WHICH PROMOTES
CC       MEMBRANE FUSION AND IS INVOLVED IN EXOCYTOSIS.
CC   -!- SIMILARITY: TO OTHER PROTEINS OF THE ANNEXIN FAMILY.
DR   EMBL; Z29443; CAA82571.1; -.
DR   HSSP; P79134; 1AVC.
DR   PROSITE; PS00223; ANNEXIN; 1.
DR   PFAM; PF00191; annexin; 4.
DR   PRINTS; PR00196; ANNEXIN.
KW   Annexin; Calcium/phospholipid-binding; Repeat.
SQ   SEQUENCE   676 AA;  74969 MW;  00E36C79 CRC32;

  Query Match             7.3%;  Score 117;  DB 5;  Length 676;
  Best Local Similarity  22.5%;  Pred. No. 1.69e-02;
  Matches    16;  Conservative   26;  Mismatches  25;  Indels   4;  Gaps   4;

Db    132 EKVALERIVEQKLAESAKNEKDIAWNVNIVVANSYVAKSLRPVINIQMPTVGGDTNFEFD 191
          ::  |: : :|  :    :  |: | |:: :: | :|:||:| : :|:  ::  : : |:
Qy    134 QRPLLDSVAQQQ-GAWLPHVADFRWRVDVAISTSALARSLQPSVLMQL-KLSDGSAYRFE 191

Db    192 IDSFAQFRQQL 202
          : : | | |:|
Qy    192 VPT-AKF-QEL 200


RESULT    4
ID   O69354      PRELIMINARY;      PRT;  1576 AA.
AC   O69354;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1999 (TrEMBLrel. 11, Last sequence update)
DT   01-AUG-1999 (TrEMBLrel. 11, Last annotation update)
DE   RECEPTOR-LIKE HISTIDINE KINASE.
GN   BPHS.
OS   Rhodococcus erythropolis.
OG   Plasmid pTA421.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TA421;
RA   ARAI H., KOSONO S., TAGUCHI K., MAEDA M., FUJI F., CHUNG S., SONG E.,
RA   KUDO T.;
RT   "Two sets of the biphenyl degradation genes on a linear plasmid of
RT   Rhodococcus erythropolis TA421.";
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 1-488 FROM N.A.
RC   STRAIN=TA421;
RX   MEDLINE; 97394954.
RA   KOSONO S., MAEDA M., FUJI F., ARAI H., KUDO T.;
RT   "Three of the seven bphC genes of Rhodococcus erythropolis TA421,
RT   isolated from a termite ecosystem, are located on an indigenous
RT   plasmid associated with biphenyl degradation.";
RL   Appl. Environ. Microbiol. 63:3282-3285(1997).
DR   EMBL; AB014348; BAA34261.1; -.
DR   EMBL; D88015; BAA25609.1; -.
KW   Receptor; Kinase; Plasmid; Hypothetical protein.
SQ   SEQUENCE   1576 AA;  171035 MW;  636B3907 CRC32;

  Query Match             7.0%;  Score 113;  DB 2;  Length 1576;
  Best Local Similarity  30.3%;  Pred. No. 5.64e-02;
  Matches    30;  Conservative   18;  Mismatches  44;  Indels   7;  Gaps   7;

Db    314 MLGDSGVGKSR-LLE-AVSEHAAQHSKVTVLRAAAFDQAPARPLGPFSSVFRDLAAH-LR 370
          :|||   :  | ||   ||   ::  :  | | :     |   ||:: :  : |  : ||
Qy     39 LLGDLDRSTFRKLLKFVVSSLQGEDCRDGVQRLGVSANLPEEQLGALLAGMHTLLQQALR 98

Db    371 THPSELQR-V-REELGELIVVVGDQVPELAGA-FGGQAP 406
            |: |     |::| ||  :  | | :||:  ||:| |
Qy     99 LPPTSLKPDTFRDQLQEL-CIPQDLVGDLASVVFGSQRP 136


RESULT    5
ID   O05201      PRELIMINARY;      PRT;  1576 AA.
AC   O05201;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-JUL-1997 (TrEMBLrel. 04, Last annotation update)
DE   RECEPTOR-LIKE HISTIDINE KINASE BPDS.
GN   BPDS.
OS   Rhodococcus sp.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=M5;
RA   LABBE D., GARNON J., LAU P.C.K.;
RL   J. Bacteriol. 179:0-0(0).
```

```
DR   EMBL; U85412; AAB52543.1; -.
SQ   SEQUENCE   1576 AA;  170944 MW;  2E087C0A CRC32;

  Query Match             7.0%;  Score 113;  DB 2;  Length 1576;
  Best Local Similarity  30.3%;  Pred. No. 5.64e-02;
  Matches    30;  Conservative   18;  Mismatches  44;  Indels   7;  Gaps   7;

Db    314 MLGDSGVGKSR-LLE-AVSEHAAQHSKVTVLRAAAFDQAPARPLGPFSSVFRDLAAH-LR 370
          :|||   :  | ||   ||   ::  :  | | :      |   ||:: :  : |  : ||
Qy     39 LLGDLDRSTFRKLLKFVVSSLQGEDCRDGVQRLGVSANLPEEQLGALLAGMHTLLQQALR 98

Db    371 THPSELQR-V-REELGELIVVVGDQVPELAGA-FGGQAP 406
            |: |     |::| ||  :  | | :||:  ||:| |
Qy     99 LPPTSLKPDTFRDQLQEL-CIPQDLVGDLASVVFGSQRP 136


RESULT    6
ID   O33934      PRELIMINARY;      PRT;   360 AA.
AC   O33934;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   ERYCII.
GN   ERYCII.
OS   Saccharopolyspora erythraea (Streptomyces erythraeus).
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC   Saccharopolyspora.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL 2338;
RX   MEDLINE; 98015410.
RA   SUMMERS R.G., DONADIO S., STAVER M.J., WENDT-PIENKOWSKI E.,
RA   HUTCHINSON C.R., KATZ L.;
RT   "Sequencing and mutagenesis of genes from the erythromycin
RT   biosynthetic gene cluster of Saccharopolyspora erythraea that are
RT   involved in L-mycarose and D-desosamine production.";
RL   Microbiology 143:0-0(0).
DR   EMBL; U77454; AAB84066.1; -.
DR   PFAM; PF00067; p450; 1.
SQ   SEQUENCE   360 AA;  38420 MW;  D5E834E6 CRC32;

  Query Match             6.7%;  Score 107;  DB 2;  Length 360;
  Best Local Similarity  25.2%;  Pred. No. 3.27e-01;
  Matches    33;  Conservative   36;  Mismatches  54;  Indels   8;  Gaps   7;

Db    150 MTAVLGAAGVLR--GAAWDARVS-LDAQLSPQQLAVTEAAVAALPADPALRAL-FAGAEM 205
          |:|| :|:  |:  | :  :||| | |||:|: :|     :: |      : | |  : :
Qy      1 MSAVGAATPYLHHPGDSHSGRVSFLGAQLPPE-VAAMARLLGDLDRSTFRKLLKFVVSSL 59

Db    206 TANTVVDAVLAVSAEPGLAE-RIADDPAAAQRTVAEVLRLHPA-LHLER-RTATAEVRLG 262
            ::   |:|  ::  : |:| :::   |: :  : : ||| |: |  :  |    |: :
Qy     60 QGEDCRDGVQRLGVSANLPEEQLGALLAGMHTLLQQALRLPPTSLKPDTFRDQLQELCIP 119

Db    263 EHVIGEGEEVV 273
          : ::|:   ||
Qy    120 QDLVGDLASVV 130


RESULT    7
ID   O54225      PRELIMINARY;      PRT;   361 AA.
AC   O54225;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   DTDP-4-KETO-6-DEOXY-HEXOSE 3,4-ISOMERASE.
GN   ERYCII.
OS   Saccharopolyspora erythraea (Streptomyces erythraeus).
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC   Saccharopolyspora.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL2338;
RA   SALAH-BEY K., DOUMITH M., MICHEL J.M., HAYDOCK S., CORTES J.,
RA   LEADLAY P.F., RAYNAL M.C.;
RL   Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Y14332; CAA74711.1; -.
DR   PFAM; PF00067; p450; 1.
KW   Isomerase.
SQ   SEQUENCE   361 AA;  38507 MW;  5FC27F6A CRC32;

  Query Match             6.7%;  Score 107;  DB 2;  Length 361;
  Best Local Similarity  25.2%;  Pred. No. 3.27e-01;
  Matches    33;  Conservative   36;  Mismatches  54;  Indels   8;  Gaps   7;

Db    151 MTAVLGAAGVLR--GAAWDARVS-LDAQLSPQQLAVTEAAVAALPADPALRAL-FAGAEM 206
          |:|| :|:  |:  | :  :||| | |||:|: :|     :: |      : | |  : :
Qy      1 MSAVGAATPYLHHPGDSHSGRVSFLGAQLPPE-VAAMARLLGDLDRSTFRKLLKFVVSSL 59

Db    207 TANTVVDAVLAVSAEPGLAE-RIADDPAAAQRTVAEVLRLHPA-LHLER-RTATAEVRLG 263
            ::   |:|  ::  : |:| :::   |: :  : : ||| |: |  :  |    |: :
Qy     60 QGEDCRDGVQRLGVSANLPEEQLGALLAGMHTLLQQALRLPPTSLKPDTFRDQLQELCIP 119

Db    264 EHVIGEGEEVV 274
          : ::|:   ||
Qy    120 QDLVGDLASVV 130


RESULT    8
ID   P72863      PRELIMINARY;      PRT;   451 AA.
AC   P72863;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   HYPOTHETICAL 49.7 KD PROTEIN.
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PCC6803;
RA   TABATA S.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PCC6803;
RX   MEDLINE; 97061201.
RA   KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA   MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA   HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
RA   SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
RA   TABATA S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
RL   DNA Res. 3:109-136(1996).
DR   EMBL; D90901; BAA16879.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   451 AA;  49667 MW;  EEC3F698 CRC32;

  Query Match             6.7%;  Score 107;  DB 2;  Length 451;
  Best Local Similarity  38.6%;  Pred. No. 3.27e-01;
  Matches    22;  Conservative   13;  Mismatches  18;  Indels   4;  Gaps   4;

Db     42 ALHFDDPRLPEVEQLTLSGT-PEE-LEALALAVETYLQQRLN-PCASLEPSTYPEPV 95
          :|: :| |   |::| :|:  ||| | ||  :: | ||| |  | :|| | |: : :
Qy     58 SLQGEDCR-DGVQRLGVSANLPEEQLGALLAGMHTLLQQALRLPPTSLKPDTFRDQL 113


RESULT    9
ID   O07346      PRELIMINARY;      PRT;   463 AA.
AC   O07346;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   PMGA,COMPLETE CDS.
```

```
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   HIHARA Y., IKEUCHI M.;
RL   Photosyn. Res. 0:0-0(0).
DR   EMBL; AB004554; BAA20428.1; -.
SQ   SEQUENCE   463 AA;  51532 MW;  944134AF CRC32;

  Query Match             6.7%;  Score 108;  DB 2;  Length 463;
  Best Local Similarity  22.3%;  Pred. No. 2.45e-01;
  Matches    31;  Conservative   41;  Mismatches  57;  Indels  10;  Gaps  10;

Db    116 RIGGFTXLPLEPLLQSPASL-GYRNKATYP-LSRSKTGQVQAGYYRKGSHRLVNINQCPV 173
          |:: |    | | : : | | |  :::|:  | :   : :|:   | | :||  ::   :
Qy     21 RVS-FLGAQLPPEVAAMARLLGDLDRSTFRKLLKFVVSSLQGEDCRDGVQRL-GVSA-NL 77

Db    174 QDDRLNLLLTEVKKDIENRGWSIYDEEKKQGKLR-HL-SLRIGQR-TGEMLLTLISAHKG 230
           :::|  ||: :   :: ::  :     |   :| :|  | | |   |::   ::::::
Qy     78 PEEQLGALLAGMHTLLQ-QALRLPPTSLKPDTFRDQLQELCIPQDLVGDLASVVFGSQRP 136

Db    231 LPD-LEEQAGEWLERYPDL 248
          | | : :| | || : :|:
Qy    137 LLDSVAQQQGAWLPHVADF 155


RESULT   10
ID   P73374      PRELIMINARY;      PRT;   463 AA.
AC   P73374;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   HYPOTHETICAL 51.5 KD PROTEIN.
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PCC6803;
RA   TABATA S.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PCC6803;
RX   MEDLINE; 97061201.
RA   KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA   MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA   HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
RA   SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
RA   TABATA S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
RL   DNA Res. 3:109-136(1996).
DR   EMBL; D90905; BAA17405.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   463 AA;  51523 MW;  7A774A5D CRC32;

  Query Match             6.7%;  Score 107;  DB 2;  Length 463;
  Best Local Similarity  22.3%;  Pred. No. 3.27e-01;
  Matches    31;  Conservative   41;  Mismatches  57;  Indels  10;  Gaps  10;

Db    116 RIGGFTTLPLEPLLQSPASL-GYRNKATYP-LSRSKTGQVQAGYYRKGSHRLVNINQCPV 173
          |:: |    | | : : | | |  :::|:  | :   : :|:   | | :||  ::   :
Qy     21 RVS-FLGAQLPPEVAAMARLLGDLDRSTFRKLLKFVVSSLQGEDCRDGVQRL-GVSA-NL 77

Db    174 QDDRLNLLLTEVKKDIENRGWSIYDEEKKQGKLR-HL-SLRIGQR-TGEMLLTLISAHKG 230
           :::|  ||: :   :: ::  :     |   :| :|  | | |   |::   ::::::
Qy     78 PEEQLGALLAGMHTLLQ-QALRLPPTSLKPDTFRDQLQELCIPQDLVGDLASVVFGSQRP 136

Db    231 LPD-LEEQAGEWLERYPDL 248
          | | : :| | || : :|:
Qy    137 LLDSVAQQQGAWLPHVADF 155


RESULT   11
ID   O43632      PRELIMINARY;      PRT;   902 AA.
AC   O43632;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   SPINDLE POLE BODY PROTEIN SPC97 HOMOLOG.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98234402.
RA   MURPHY S.M., URBANI L., STEARNS T.;
RT   "The mammalian gamma-tubulin complex contains homologues of the yeast
RT   spindle pole body components spc97p and spc98p.";
RL   J. Cell Biol. 141:663-674(1998).
DR   EMBL; AF042379; AAC39728.1; -.
SQ   SEQUENCE   902 AA;  102533 MW;  A4ED2937 CRC32;

  Query Match             6.6%;  Score 106;  DB 4;  Length 902;
  Best Local Similarity  37.2%;  Pred. No. 4.36e-01;
  Matches    16;  Conservative   15;  Mismatches   8;  Indels   4;  Gaps   4;

Db    551 LEALLELALRMS-TA-NTDPFKDDLKIDL-MPHDLITQLLRVL 590
          : :||: |||:: |: : | |:|:|  :| :|:||: :|  |:
Qy     89 MHTLLQQALRLPPTSLKPDTFRDQLQ-ELCIPQDLVGDLASVV 130


RESULT   12
ID   O96777      PRELIMINARY;      PRT;   678 AA.
AC   O96777;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   CYCLIC NUCLEOTIDE AND VOLTAGE-ACTIVATED ION CHANNEL.
GN   CNG.
OS   Heliothis virescens (Noctuid moth) (Owlet moth).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Noctuoidea; Noctuidae; Heliothinae; Heliothis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=ANTENNAE;
RA   KRIEGER J., STROBEL J., VOGL A., HANKE W., BREER H.;
RT   "Identification of a cyclic nucleotide and voltage-activated ion
RT   channel from insect antennae.";
RL   Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ012664; CAA10110.1; -.
DR   HSSP; P00514; 1APK.
SQ   SEQUENCE   678 AA;  76305 MW;  CD1BDE7A CRC32;

  Query Match             6.5%;  Score 105;  DB 5;  Length 678;
  Best Local Similarity  36.7%;  Pred. No. 5.81e-01;
  Matches    22;  Conservative   11;  Mismatches  22;  Indels   5;  Gaps   4;

Db     59 PATDAKQGFLRNQLQALFQPTD--NKLAMKLFGSKKALMKERIRQKAAGHWVIHPCSSFR 116
          |:|  |   :|:|||  |  | |    ||  :||| ::|: : : |   | |:|  : ||
Qy    100 PPTSLKPDTFRDQLQELCIPQDLVGDLASVVFGSQRPLL-DSVAQQQ-GAWLPHV-ADFR 156


RESULT   13
ID   Q9WV07      PRELIMINARY;      PRT;   711 AA.
AC   Q9WV07;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   LIPOXYGENASE-3.
GN   ALOXE3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NMRI; TISSUE=SKIN EPIDERMIS: SQUAMOUS CELL CARCINOMA;
RX   MEDLINE; 99296830.
RA   KINZIG A., HEIDT M., FUERSTENBERGER G., MARKS F., KRIEG P.;
RT   "cDNA cloning, genomic structure, and chromosomal localization of a
RT   novel murine epidermis-type lipoxygenase.";
RL   Genomics 58:158-164(1999).
DR   EMBL; Y14695; CAB46101.1; -.
SQ   SEQUENCE   711 AA;  80578 MW;  B9BFE292 CRC32;

  Query Match            6.5%;  Score 104;  DB 11;  Length 711;
  Best Local Similarity  35.0%;  Pred. No. 7.71e-01;
  Matches    14;  Conservative   14;  Mismatches  10;  Indels   2;  Gaps   2;

Db    567 LCTPGELVKYLTAIIFNCSAQHAAFNSGQHDFGAWMPNAP 606
          || | :||  |::::|  | |:: ::|  :: |||:|: :
Qy    116 LCIPQDLVGDLASVVFG-S-QRPLLDSVAQQQGAWLPHVA 153


RESULT   14
ID   Q9ZBT5      PRELIMINARY;      PRT;   327 AA.
AC   Q9ZBT5;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   PUTATIVE REGULATORY PROTEIN.
GN   SC1A9.17C.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   SAUNDERS D.C., HARRIS D.;
RL   Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   BENTLEY S.D., PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL   Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE; 97000351.
RA   REDENBACH M., KIESER H.M., DENAPAITE D., EICHNER A., CULLUM J.,
RA   KINASHI H., HOPWOOD D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
DR   EMBL; AL034446; CAA22387.1; -.
SQ   SEQUENCE   327 AA;  35032 MW;  A2BC81BD CRC32;

  Query Match            6.4%;  Score 103;  DB 2;  Length 327;
  Best Local Similarity  28.5%;  Pred. No. 1.02e+00;
  Matches    35;  Conservative   33;  Mismatches  47;  Indels   8;  Gaps   7;

Db      9 DTHEAAYRALVSVGAADVPDLARRLALGERDTERALRRLEQNGLAAQSSARPG--RW-VA 65
          |:| :    | :  :::|: :|| |:  :| | | | ::  ::| ::   | |  |  |:
Qy     16 DSHSGRVSFLGAQLPPEVAAMARLLGDLDRSTFRKLLKFVVSSLQGEDC-RDGVQRLGVS 74

Db     66 AP-PGVALGALLTQQRHELERA-ELA-AALLAEEYRAAAAEPAV-HDLVEVVTGAGAVAQ 121
          |  |   |||||:  :  |::|  |: ::| :: :|    |  : :|||  :::    :|
Qy     75 ANLPEEQLGALLAGMHTLLQQALRLPPTSLKPDTFRDQLQELCIPQDLVGDLASVVFGSQ 134

Db    122 RFL 124
          | |
Qy    135 RPL 137


RESULT   15
ID   O53544      PRELIMINARY;      PRT;   400 AA.
AC   O53544;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   MCE-FAMILY PROTEIN.
GN   MTV023.06C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RA   BADCOCK K., CHURCHER C.M.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RA   COLE S.T., PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE; 96181548.
RA   PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,
RA   BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
RA   COLE S.T.;
RT   "An integrated map of the genome of the tubercle bacillus,
RT   Mycobacterium tuberculosis H37Rv, and comparison with Mycobacterium
RT   leprae.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:3132-3137(1996).
DR   EMBL; AL022022; CAA17736.1; -.
SQ   SEQUENCE   400 AA;  42418 MW;  8C21A90C CRC32;

  Query Match            6.4%;  Score 102;  DB 2;  Length 400;
  Best Local Similarity  26.3%;  Pred. No. 1.35e+00;
  Matches    31;  Conservative   31;  Mismatches  48;  Indels   8;  Gaps   7;

Db    126 LSPNAHVAASQV-QLEVNTLFQSLIDLL-HKIDPLETNATLSALSEGLRGHGDDLGALLS 183
          |:|:    |  : :|: :| |: |: ::   ::  :    :  |: :     ::|||||:
Qy     29 LPPEVAAMARLLGDLDRST-FRKLLKFVVSSLQGEDCRDGVQRLGVSANLPEEQLGALLA 87

Db    184 GLNTLTRQANPKLP--ALQED-FRKAAVVANVYADAAGDLNTV-FDNLPTINKTIVDQ 237
          |::|| :|| :|| :|  | ||       :  |  ||| :| | :  :  :: :|
Qy     88 GMHTLLQQA-LRLPPTSLKPDTFRDQLQELCIPQDLVGDLASVVFGSQRPLLDSVAQQ 144


Search completed: Fri May 12 10:33:54 2000
Job time : 87 secs.
```

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 12 10:31:30 2000; MasPar time 34.81 Seconds
192.185 Million cell updates/sec
Tabular output not generated.

Title: >US-09-223-796-4
Description: (1-224) from US09223796.pep
Perfect Score: 1604
Sequence: 1 MSAVGAATPYLHHPGDSHSG.....ALVLKEMADLEKRCERRLQD 224

Scoring table: PAM 150
Gap 11

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 47.835; Variance 94.148; scale 0.508

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	110	6.9	2414	1	P300_HUMAN	E1A-ASSOCIATED PROTEIN	5.73e-02
2	109	6.8	234	1	NOLW_RHIFR	NODULATION PROTEIN NOL	7.74e-02
3	109	6.8	234	1	NOLW_RHISN	NODULATION PROTEIN NOL	7.74e-02
4	108	6.7	230	1	HEMX_HAEIN	PUTATIVE UROPORPHYRIN-	1.04e-01
5	108	6.7	282	1	LIPH_PSEAE	LIPH PROTEIN.	1.04e-01
6	108	6.7	302	1	VPJ_BPP2	BASEPLATE ASSEMBLY PRO	1.04e-01
7	106	6.6	340	1	LIMA_PSEAE	LIPASE MODULATOR PRECU	1.89e-01
8	103	6.4	2670	1	YAQ5_SCHPO	PUTATIVE TRANSLATIONAL	4.54e-01
9	102	6.4	3255	1	POLG_LMVE	GENOME POLYPROTEIN [CO	6.05e-01
10	101	6.3	81	1	RUXX_METTH	PUTATIVE SNRNP SM-LIKE	8.05e-01
11	99	6.2	331	1	SYFA_SYNY3	PHENYLALANYL-TRNA SYNT	1.42e+00
12	99	6.2	570	1	FLIF_RHOSH	FLAGELLAR M-RING PROTE	1.42e+00
13	100	6.2	784	1	GCF_HUMAN	GC-RICH SEQUENCE DNA-B	1.07e+00
14	100	6.2	946	1	YIN7_YEAST	HYPOTHETICAL ZINC AMIN	1.07e+00
15	100	6.2	1144	1	RIR1_HSV23	RIBONUCLEOSIDE-DIPHOSP	1.07e+00
16	98	6.1	434	1	YADA_YERPS	INVASIN PRECURSOR (OUT	1.87e+00
17	97	6.0	326	1	GSPK_KLEPN	GENERAL SECRETION PATH	2.47e+00
18	97	6.0	661	1	TRA1_HUMAN	TUMOR NECROSIS FACTOR	2.47e+00
19	95	5.9	189	1	AROK_SYNY3	SHIKIMATE KINASE (EC 2	4.27e+00
20	94	5.9	306	1	YLIC_ECOLI	HYPOTHETICAL ABC TRANS	5.60e+00
21	94	5.9	348	1	HOXV_AZOVI	HYDROGENASE EXPRESSION	5.60e+00
22	94	5.9	499	1	CPDF_CANFA	CYTOCHROME P450 2D15 (	5.60e+00
23	95	5.9	603	1	ALS_RAT	INSULIN-LIKE GROWTH FA	4.27e+00
24	94	5.9	1137	1	RIR1_HSV11	RIBONUCLEOSIDE-DIPHOSP	5.60e+00
25	95	5.9	2672	1	GCN1_YEAST	TRANSLATIONAL ACTIVATO	4.27e+00
26	93	5.8	336	1	YDBI_SCHPO	HYPOTHETICAL 38.4 KD P	7.32e+00
27	93	5.8	560	1	YM19_YEAST	HYPOTHETICAL 63.1 KD P	7.32e+00
28	93	5.8	600	1	LAM2_CHICK	LAMIN B2.	7.32e+00
29	93	5.8	603	1	ALS_MOUSE	INSULIN-LIKE GROWTH FA	7.32e+00
30	93	5.8	692	1	YZN4_CAEEL	HYPOTHETICAL ACETYLCHO	7.32e+00
31	93	5.8	836	1	VG26_BPMD2	MINOR TAIL PROTEIN GP2	7.32e+00
32	93	5.8	842	1	PKL2_RAT	PROTEIN KINASE C-LIKE	7.32e+00
33	93	5.8	886	1	YFIQ_ECOLI	HYPOTHETICAL 98.0 KD P	7.32e+00
34	93	5.8	1276	1	GIL1_ENTHI	GALACTOSE-INHIBITABLE	7.32e+00
35	93	5.8	3255	1	POLG_LMV0	GENOME POLYPROTEIN [CO	7.32e+00
36	91	5.7	131	1	DFRA_MYXXA	DFRA PROTEIN.	1.24e+01
37	91	5.7	162	1	IL2_CEREL	INTERLEUKIN-2 PRECURSO	1.24e+01
38	92	5.7	212	1	Y010_MYCPN	HYPOTHETICAL PROTEIN M	9.54e+00
39	91	5.7	280	1	LBX1_MOUSE	TRANSCRIPTION FACTOR L	1.24e+01
40	91	5.7	393	1	YJGN_ECOLI	HYPOTHETICAL 44.4 KD P	1.24e+01
41	92	5.7	538	1	VGLF_MUMPM	FUSION GLYCOPROTEIN PR	9.54e+00
42	92	5.7	558	1	YAB3_SCHPO	PUTATIVE VACUOLAR PROT	9.54e+00
43	91	5.7	603	1	ENV_RSVP	ENV POLYPROTEIN [CONTA	1.24e+01
44	91	5.7	753	1	YA79_SCHPO	HYPOTHETICAL 84.5 KD P	1.24e+01
45	92	5.7	2329	1	YS89_CAEEL	HYPOTHETICAL 254.3 KD	9.54e+00

ALIGNMENTS

```
RESULT    1
ID   P300_HUMAN     STANDARD;      PRT;  2414 AA.
AC   Q09472;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   E1A-ASSOCIATED PROTEIN P300.
GN   EP300 OR P300.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95011587.
RA   ECKNER R., EWEN M.E., NEWSOME D., GERDES M., DECAPRIO J.A.,
RA   LAWRENCE J.B., LIVINGSTON D.M.;
RT   "Molecular cloning and functional analysis of the adenovirus E1A-
RT   associated 300-kD protein (p300) reveals a protein with properties of
RT   a transcriptional adaptor.";
RL   Genes Dev. 8:869-884(1994).
CC   -!- FUNCTION: PROBABLE TRANSCRIPTIONAL ADAPTOR REQUIRED FOR THE
CC       ACTIVITY OF CERTAIN COMPLEX TRANSCRIPTIONAL REGULATORY ELEMENTS.
CC       MAY HAVE A FUNCTION IN CELL CYCLE REGULATION. BINDS TO AND MAY BE
CC       INVOLVED IN THE TRANSFORMING CAPACITY OF THE ADENOVIRUS E1A
CC       PROTEIN.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U01877; AAA18639.1; -.
DR   MIM; 602700; -.
DR   PROSITE; PS00633; BROMODOMAIN_1; 1.
DR   PROSITE; PS50014; BROMODOMAIN_2; 1.
DR   PFAM; PF00439; bromodomain; 1.
DR   PFAM; PF00569; ZZ; 1.
KW   Transcription regulation; Nuclear protein; Bromodomain; Cell cycle.
FT   DOMAIN       11     17       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN      797    800       POLY-SER.
FT   DOMAIN     1067   1139       BROMODOMAIN.
FT   DOMAIN     1519   1526       POLY-GLU.
```

```
FT   DOMAIN      1572   1818       BINDING REGION FOR E1A ADENOVIRUS.
FT   DOMAIN      2066   2069       POLY-GLN.
FT   DOMAIN      2190   2195       POLY-GLN.
SQ   SEQUENCE   2414 AA;  264143 MW;  ED6169CC CRC32;

  Query Match             6.9%;  Score 110;  DB 1;  Length 2414;
  Best Local Similarity  34.0%;  Pred. No. 5.73e-02;
  Matches    17;  Conservative   13;  Mismatches  18;  Indels   2;  Gaps   2;

Db   2137 QAGVQRAGLPQQQPQQQLQPPMGGMSPQAQQ-MNMNHNTM-PSQFRDILR 2184
            : |||| |::   |::|| : ::||    || : :   :: |  ||| |:
Qy     65 RDGVQRLGVSANLPEEQLGALLAGMHTLLQQALRLPPTSLKPDTFRDQLQ 114


RESULT    2
ID   NOLW_RHIFR     STANDARD;      PRT;   234 AA.
AC   P33212;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   NODULATION PROTEIN NOLW.
GN   NOLW.
OS   Rhizobium fredii.
OG   Plasmid sym.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Sinorhizobium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=USDA 257;
RX   MEDLINE; 94018604.
RA   MEINHARDT L.W., KRISHNAN H.B., BALATTI P.A., PUEPPKE S.G.;
RT   "Molecular cloning and characterization of a sym plasmid locus that
RT   regulates cultivar-specific nodulation of soybean by Rhizobium fredii
RT   USDA257.";
RL   Mol. Microbiol. 9:17-29(1993).
CC   -!- FUNCTION: REGULATES CULTIVAR-SPECIFIC NODULATION OF SOYBEAN.
CC   -!- SIMILARITY: SOME, TO THE PULD/OUTD/EXED/XPSD FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L12251; AAB17675.1; -.
DR   PIR; S35020; S35020.
KW   Plasmid; Nodulation; Transmembrane.
FT   TRANSMEM     17     36       POTENTIAL.
SQ   SEQUENCE   234 AA;  25829 MW;  87C798A5 CRC32;

  Query Match             6.8%;  Score 109;  DB 1;  Length 234;
  Best Local Similarity  27.0%;  Pred. No. 7.74e-02;
  Matches    34;  Conservative   35;  Mismatches  49;  Indels   8;  Gaps   8;

Db     24 LFAGIHTTLGATLPLPSTSYKYTVLDQDLSAALQEFGNNLKISVNISAEVKGRIRGRIPE 83
          |:||:|| |  :| ||:|| |   : ::|   |  :  :| :: :::: | |  |  :
Qy     85 LLAGMHTLLQQALRLPPTSLKPDTFRDQL-QELC-IPQDL-VG-DLASVVFGSQRPLLDS 140

Db     84 LSPREFLDRLTDLYDLQWYYDGVVLYVSA-AKEAQTRMLV-LSSVHFSAFKLALDKLDIS 141
          :: ::   :|  : |::|  | | :  || |:  |  :|: |     ||::: :    :
Qy    141 VAQQQG-AWLPHVADFRWRVD-VAISTSALARSLQPSVLMQLKLSDGSAYRFEVPTAKFQ 198

Db    142 DERYPV 147
          : ||:|
Qy    199 ELRYSV 204


RESULT    3
ID   NOLW_RHISN     STANDARD;      PRT;   234 AA.
AC   P55712;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   NODULATION PROTEIN NOLW.
GN   NOLW OR Y4YD.
OS   Rhizobium sp. (strain NGR234).
OG   Plasmid sym pNGR234a.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Rhizobium.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97305956.
RA   FREIBERG C.A., FELLAY R., BAIROCH A., BROUGHTON W.J., ROSENTHAL A.,
RA   PERRET X.;
RT   "Molecular basis of symbiosis between Rhizobium and legumes.";
RL   Nature 387:394-401(1997).
CC   -!- SIMILARITY: SOME, TO THE PULD/OUTD/EXED/XPSD FAMILY.
CC   -!- SIMILARITY: SOME, TO Y4XJ.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000107; AAB91943.1; -.
KW   Nodulation; Plasmid; Transmembrane.
FT   TRANSMEM     17     38       POTENTIAL.
SQ   SEQUENCE   234 AA;  25847 MW;  392E7F48 CRC32;

  Query Match             6.8%;  Score 109;  DB 1;  Length 234;
  Best Local Similarity  27.0%;  Pred. No. 7.74e-02;
  Matches    34;  Conservative   35;  Mismatches  49;  Indels   8;  Gaps   8;

Db     24 LFAGIHTTLGATLPLPSTSYKYTVLDQDLSAALQEFGNNLKISVNISAEVKGRIRGRIPE 83
          |:||:|| |  :| ||:|| |   : ::|   |  :  :| :: :::: | |  |  :
Qy     85 LLAGMHTLLQQALRLPPTSLKPDTFRDQL-QELC-IPQDL-VG-DLASVVFGSQRPLLDS 140

Db     84 LSPREFLDRLTDLYDLQWYYDGVVLYVSA-AKEAQTRMLV-LSSVHFSAFKLALDKLDIS 141
          :: ::   :|  : |::|  | | :  || |:  |  :|: |     ||::: :    :
Qy    141 VAQQQG-AWLPHVADFRWRVD-VAISTSALARSLQPSVLMQLKLSDGSAYRFEVPTAKFQ 198

Db    142 DERYPV 147
          : ||:|
Qy    199 ELRYSV 204


RESULT    4
ID   HEMX_HAEIN     STANDARD;      PRT;   230 AA.
AC   P44773;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   PUTATIVE UROPORPHYRIN-III C-METHYLTRANSFERASE (EC 2.1.1.107) (UROGEN
DE   III METHYLASE).
GN   HEMX OR HI0603.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RD / KW20;
RX   MEDLINE; 95350630.
RA   FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
RA   KERLAVAGE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,
RA   MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,
RA   SCOTT J.D., SHIRLEY R., LIU L.-I., GLODEK A., KELLEY J.M.,
RA   WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOM E., COTTON M.D.,
RA   UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,
RA   FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,
RA   GNEHM C.L., MCDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
RA   VENTER J.C.;
```

```
RT   "Whole-genome random sequencing and assembly of Haemophilus
RT   influenzae Rd.";
RL   Science 269:496-512(1995).
CC   -!- CATALYTIC ACTIVITY: 2 S-ADENOSYL-L-METHIONINE + UROPORPHYRIN III
CC       = 2 S-ADENOSYL-L-HOMOCYSTEINE + SIROHYDROCHLORIN.
CC   -!- PATHWAY: INVOLVED IN THE BIOSYNTHESIS OF SIROHEME AND COBALAMIN.
CC   -!- SIMILARITY: MODERATE SIMILARITY TO E.COLI HEMX.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32742; AAC22261.1; -.
DR   TIGR; HI0603; -.
KW   Porphyrin biosynthesis; Transferase; Methyltransferase.
SQ   SEQUENCE   230 AA;  25193 MW;  432868A7 CRC32;

  Query Match             6.7%;  Score 108;  DB 1;  Length 230;
  Best Local Similarity  27.5%;  Pred. No. 1.04e-01;
  Matches    38;  Conservative   34;  Mismatches  56;  Indels  10;  Gaps   7;

Db     76 ENQTGANLSSNNTNNNKRLTQLEQSLKTAQENIAQLEQLIVSK--TGEITSLQTQMKQVS 133
          |:| || |:: :|     |     ||| ::    ||::| ::    |:::|:     | :
Qy     79 EEQLGALLAGMHTLLQQALRLPPTSLK-PDTFRDQLQELCIPQDLVGDLASVVFG-SQRP 136

Db    134 QL-AIAQQPSDWLFSEADFL--LNNALRKLVLDNDVDTAVSL-LKLADETLVK--VNNSQ 187
           | ::||| : ||   |||   :: |:    |   :: :| : |||:| :  :  |  :
Qy    137 LLDSVAQQQGAWLPHVADFRWRVDVAISTSALARSLQPSVLMQLKLSDGSAYRFEVPTAK 196

Db    188 ANEIRSAINQDLKQLLSL 205
            |:| ::   ||::  |
Qy    197 FQELRYSVALVLKEMADL 214


RESULT    5
ID   LIPH_PSEAE     STANDARD;      PRT;   282 AA.
AC   Q01725;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   LIPH PROTEIN.
GN   LIPH.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC   Pseudomonas.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE; 92381478.
RA   WOHLFARTH S., HOESCHE C., STRUNK C., WINKLER U.K.;
RT   "Molecular genetics of the extracellular lipase of Pseudomonas
RT   aeruginosa PAO1.";
RL   J. Gen. Microbiol. 138:1325-1335(1992).
RN   [2]
RP   REVISION TO 244-274.
RC   STRAIN=ATCC 15692 / PAO1;
RA   WOHLFARTH S.;
RL   Submitted (MAY-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: HELPS TO PRODUCE AN EXTRACELLULARLY ACTIVE LIPASE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X63391; CAA44998.1; -.
DR   PIR; S25769; S25769.
SQ   SEQUENCE   282 AA;  31816 MW;  3252D3E5 CRC32;

  Query Match             6.7%;  Score 108;  DB 1;  Length 282;
  Best Local Similarity  30.5%;  Pred. No. 1.04e-01;
  Matches    18;  Conservative   22;  Mismatches  16;  Indels   3;  Gaps   2;

Db    150 LSAEEKAAAIDRL--RASLPEDQQESVLPQLQSELQQQTAALQAAGAGPEAIRQMRQQL 206
          | :|:   :::||   |:|||:|  ::|: ::: | ||:  | :::  |:::|:  |:|
Qy     59 LQGEDCRDGVQRLGVSANLPEEQLGALLAGMHT-LLQQALRLPPTSLKPDTFRDQLQEL 116


RESULT    6
ID   VPJ_BPP2       STANDARD;      PRT;   302 AA.
AC   P51767;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-DEC-1999 (Rel. 39, Last annotation update)
DE   BASEPLATE ASSEMBLY PROTEIN J (GPJ).
GN   J.
OS   Bacteriophage P2.
OC   Viruses; dsDNA viruses, no RNA stage; Tailed phages; Myoviridae.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96036485.
RA   HAGGARD-LJUNGQUIST E., JACOBSEN E., RISHOVD S., SIX E.W., NILSSEN O.,
RA   SUNSHINE M.G., LINDQVIST B.H., KIM K.-J., BARREIRO V., KOONIN E.V.,
RA   CALENDAR R.;
RT   "Bacteriophage P2: genes involved in baseplate assembly.";
RL   Virology 213:109-121(1995).
CC   -!- FUNCTION: LIES AT THE EDGE OF THE BASEPLATE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF063097; AAD03284.1; -.
SQ   SEQUENCE   302 AA;  32780 MW;  CB8BBA88 CRC32;

  Query Match             6.7%;  Score 108;  DB 1;  Length 302;
  Best Local Similarity  31.8%;  Pred. No. 1.04e-01;
  Matches    28;  Conservative   18;  Mismatches  39;  Indels   3;  Gaps   3;

Db    135 LSVAGSVGAYQYH-GRSADGRVADISVTSPSPACVTISVLSRENNGVASEDLLAVVRNAL 193
          :|  |:  :|  | | |  |||: ::   | |   ::: |  : :  :   ||  | ::|
Qy      1 MSAVGAATPYLHHPGDSHSGRVSFLGAQLP-PEVAAMARLLGDLDRSTFRKLLKFVVSSL 59

Db    194 NGEDVRPVADRVTVQSAAIVEYQINATL 221
            ||| |   :|: | || : | |: | |
Qy     60 QGEDCRDGVQRLGV-SANLPEEQLGALL 86


RESULT    7
ID   LIMA_PSEAE     STANDARD;      PRT;   340 AA.
AC   Q04591; P95420;
DT   01-OCT-1993 (Rel. 27, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   LIPASE MODULATOR PRECURSOR (LIPASE HELPER PROTEIN).
GN   LIPB.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC   Pseudomonas.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TE3285;
RX   MEDLINE; 92337414.
RA   CHIHARA-SIOMI M., YOSHIKAWA K., OSHIMA-HIRAYAMA N., YAMAMOTO K.,
RA   SOGABE Y., NAKATANI T., NISHIOKA T., ODA J.;
```

```
RT   "Purification, molecular cloning, and expression of lipase from
RT   Pseudomonas aeruginosa.";
RL   Arch. Biochem. Biophys. 296:505-513(1992).
RN   [2]
RP   REVISIONS.
RC   STRAIN=TE3285;
RA   SHIBATA H.;
RL   Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PARTICIPATES IN LIPASE PRODUCTION. THE LIPASE GENE IS
CC       WELL EXPRESSED ONLY IN THE PRESENCE OF THIS PROTEIN WHICH EXERTS
CC       ITS EFFECT BOTH IN CIS AND IN TRANS. IT MAY IN SOME WAY
CC       PARTICIPATE IN THE SECRETION OF THE LIPASE (BY SIMILARITY).
CC   -!- SIMILARITY: HIGH, WITH OTHER MODULATORS OF THE PRODUCTION OF
CC       LIPASES IN OTHER PSEUDOMONAS SPECIES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB008452; BAA23129.1; -.
DR   PIR; S24161; S24161.
KW   Lipid degradation; Signal.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18    340       LIPASE MODULATOR.
SQ   SEQUENCE   340 AA;  37686 MW;  571A5FAC CRC32;

  Query Match           6.6%;  Score 106;  DB 1;  Length 340;
  Best Local Similarity  30.5%;  Pred. No. 1.89e-01;
  Matches   18;  Conservative   21;  Mismatches  17;  Indels   3;  Gaps   2;

Db    208 LSTEEKAAAIDRL--RASLPEDQQESVLPQLQSELQQQTAALQAAGAGPEAIRQMRQQL 264
          |  |:   :::||   |:|||:|  ::|: ::: | ||:  | :::  |:::|:  |:|
Qy     59 LQGEDCRDGVQRLGVSANLPEEQLGALLAGMHT-LLQQALRLPPTSLKPDTFRDQLQEL 116


RESULT    8
ID   YAQ5_SCHPO     STANDARD;      PRT;  2670 AA.
AC   Q10105;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   PUTATIVE TRANSLATIONAL ACTIVATOR C18G6.05C (GCN1 HOMOLOG).
GN   SPAC18G6.05C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Archiascomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   CONNOR R., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL   Submitted (DEC-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: STRONG, TO YEAST GCN1.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z68198; CAA92385.1; -.
KW   Hypothetical protein; Translation regulation; Activator.
SQ   SEQUENCE   2670 AA;  297333 MW;  5C214AA7 CRC32;

  Query Match           6.4%;  Score 103;  DB 1;  Length 2670;
  Best Local Similarity  26.5%;  Pred. No. 4.54e-01;
  Matches   22;  Conservative   25;  Mismatches  30;  Indels   6;  Gaps   6;

Db   1664 ALGSLIEKLGEKKFPTLIPELFNVLRSE-CSE-VDRQGAAQGLSEI-LAGLGLARLEDVL 1720
          |:: |:  |    |  |:  : : |::| | : |:| | :  |:|  |::| || :  :|
Qy     35 AMARLLGDLDRSTFRKLLKFVVSSLQGEDCRDGVQRLGVSANLPEEQLGAL-LAGMHTLL 93

Db   1721 PEILKNTSSPV-PH-IRESFISL 1741
           : |:  :::: |  :|: :  |
Qy     94 QQALRLPPTSLKPDTFRDQLQEL 116


RESULT    9
ID   POLG_LMVE      STANDARD;      PRT;  3255 AA.
AC   P89876;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   GENOME POLYPROTEIN [CONTAINS: N-TERMINAL PROTEIN (P1); HELPER
DE   COMPONENT PROTEINASE (EC 3.4.22.-) (HC-PRO); PROTEIN P3; 6 KD PROTEIN
DE   1 (6K1); CYTOPLASMIC INCLUSION PROTEIN (CI); 6 KD PROTEIN 2 (6K2);
DE   GENOME-LINKED PROTEIN (VPG); NUCLEAR INCLUSION PROTEIN A (NI-A) (NIA)
DE   (EC 3.4.22.-) (49 KD PROTEINASE) (49 KD-PRO); NUCLEAR INCLUSION
DE   PROTEIN B (NI-B) (NIB) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48);
DE   COAT PROTEIN (CP)].
OS   Lettuce mosaic virus (strain E) (LMV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Potyviridae;
OC   Potyvirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97239892.
RA   REVERS F., YANG S.J., WALTER J., SOUCHE S., LOT H., LE GALL O.,
RA   CANDRESSE T., DUNEZ J.;
RT   "Comparison of the complete nucleotide sequences of two isolates of
RT   lettuce mosaic virus differing in their biological properties.";
RL   Virus Res. 47:167-177(1997).
CC   -!- FUNCTION: HELPER COMPONENT-PROTEINASE IS REQUIRED FOR APHID
CC       TRANSMISSION AND ALSO HAS PROTEOLYTIC ACTIVITY.
CC   -!- FUNCTION: CYTOPLASMIC INCLUSION PROTEIN HAS HELICASE ACTIVITY. IT
CC       MAY BE INVOLVED IN REPLICATION.
CC   -!- FUNCTION: NUCLEAR INCLUSION PROTEIN A HAS PROTEOLYTIC ACTIVITY.
CC   -!- PTM: VPG IS COVALENTLY LINKED TO THE GENOMIC RNA.
CC   -!- PTM: THE VIRAL RNA OF POTYVIRUSES IS EXPRESSED AS A SINGLE
CC       POLYPROTEIN WHICH UNDERGOES POSTTRANSLATIONAL PROTEOLYTIC
CC       PROCESSING RESULTING IN THE PRODUCTION OF AT LEAST EIGHT
CC       INDIVIDUAL PROTEINS.
CC   -!- SIMILARITY: HC PROTEINASE BELONGS TO PEPTIDASE FAMILY C6.
CC   -!- SIMILARITY: NI-A PROTEINASE BELONGS TO PEPTIDASE FAMILY C4.
CC   -!- SIMILARITY: BELONGS TO THE POTYVIRUSES POLYPROTEIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X97705; CAA66281.1; -.
DR   PFAM; PF00270; DEAD; 1.
DR   PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR   PFAM; PF00767; Poty_coat; 1.
DR   PFAM; PF00851; Peptidase_C6; 1.
DR   PFAM; PF00863; Peptidase_C4; 1.
DR   PFAM; PF01577; Poty_P1; 1.
KW   Hydrolase; Transferase; Thiol protease; RNA-directed RNA polymerase;
KW   Coat protein; Polyprotein; Covalent protein-RNA linkage; Helicase;
KW   ATP-binding.
FT   CHAIN         1      ?       N-TERMINAL PROTEIN.
FT   CHAIN         ?      ?       HELPER COMPONENT PROTEINASE.
FT   CHAIN         ?      ?       PROTEIN P3.
FT   CHAIN         ?      ?       6 KD PROTEIN 1.
FT   CHAIN         ?      ?       CYTOPLASMIC INCLUSION PROTEIN.
FT   CHAIN         ?      ?       6 KD PROTEIN 2.
FT   CHAIN         ?      ?       GENOME-LINKED PROTEIN.
```

```
FT   CHAIN         ?      ?       NUCLEAR INCLUSION PROTEIN A.
FT   CHAIN         ?   2977       NUCLEAR INCLUSION PROTEIN B.
FT   CHAIN      2978   3255       COAT PROTEIN.
FT   NP_BIND    1410   1417       ATP (POTENTIAL).
SQ   SEQUENCE   3255 AA;  367621 MW;  A5475B9E CRC32;

  Query Match             6.4%;  Score 102;  DB 1;  Length 3255;
  Best Local Similarity  27.8%;  Pred. No. 6.05e-01;
  Matches    20;  Conservative   14;  Mismatches  32;  Indels   6;  Gaps   6;

Db   1156 WVGVQVQGVKWLAKSVHYMIPELTNILNVGTLL-LTLISLGV-RFRSLTGQFKEMKYKET 1213
           |:  :|   :|    |    : |:  |  : |: | |   :  ||   |: | |::|  :
Qy    148 WLP-HVADFRWRV-DVAISTSALARSLQPSVLMQLKLSDGSAYRFEVPTAKFQELRYSVA 205

Db   1214 LA-RE-EELRKR 1223
           |  :|  :| ||
Qy    206 LVLKEMADLEKR 217


RESULT   10
ID   RUXX_METTH     STANDARD;      PRT;    81 AA.
AC   O26745;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   PUTATIVE SNRNP SM-LIKE PROTEIN.
GN   MTH649.
OS   Methanobacterium thermoautotrophicum.
OC   Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC   Methanobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DELTA H;
RX   MEDLINE; 98037514.
RA   SMITH D.R., DOUCETTE-STAMM L.A., DELOUGHERY C., LEE H.-M., DUBOIS J.,
RA   ALDREDGE T., BASHIRZADEH R., BLAKELY D., COOK R., GILBERT K.,
RA   HARRISON D., HOANG L., KEAGLE P., LUMM W., POTHIER B., QIU D.,
RA   SPADAFORA R., VICARE R., WANG Y., WIERZBOWSKI J., GIBSON R.,
RA   JIWANI N., CARUSO A., BUSH D., SAFER H., PATWELL D., PRABHAKAR S.,
RA   MCDOUGALL S., SHIMER G., GOYAL A., PIETROVSKI S., CHURCH G.M.,
RA   DANIELS C.J., MAO J.-I., RICE P., NOLLING J., REEVE J.N.;
RT   "Complete genome sequence of Methanobacterium thermoautotrophicum
RT   deltaH: functional analysis and comparative genomics.";
RL   J. Bacteriol. 179:7135-7155(1997).
CC   -!- SIMILARITY: BELONGS TO THE SNRNP SM PROTEINS FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE000845; AAB85154.1; -.
DR   PFAM; PF01423; Sm; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   81 AA;  9061 MW;  60D0F649 CRC32;

  Query Match             6.3%;  Score 101;  DB 1;  Length 81;
  Best Local Similarity  34.8%;  Pred. No. 8.05e-01;
  Matches    23;  Conservative   16;  Mismatches  19;  Indels   8;  Gaps   7;

Db      8 RVNVQRPLDALGNSLNSPVIIKLK-GD-REFR-GVLKS-F-DLH--MNLVLNDAEELEDG 60
          ||:|  :  ||: || ::|:: || :|   :|  |  : | :|:  : |||::  :||
Qy    158 RVDVAISTSALARSLQPSVLMQLKLSDGSAYRFEVPTAKFQELRYSVALVLKEMADLEK- 216

Db     61 EVTRRL 66
             |||
Qy    217 RCERRL 222


RESULT   11
ID   SYFA_SYNY3     STANDARD;      PRT;   331 AA.
AC   Q55187;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   PHENYLALANYL-TRNA SYNTHETASE ALPHA CHAIN (EC 6.1.1.20) (PHENYLALANINE-
DE   -TRNA LIGASE ALPHA CHAIN) (PHERS).
GN   PHES OR SLL0454.
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96127529.
RA   KANEKO T., TANAKA A., SATO S., KOTANI H., SAZUKA T., MIYAJIMA N.,
RA   SUGIURA M., TABATA S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT   region from map positions 64% to 92% of the genome.";
RL   DNA Res. 2:153-166(1995).
CC   -!- CATALYTIC ACTIVITY: ATP + L-PHENYLALANINE + TRNA(PHE) = AMP +
CC       PYROPHOSPHATE + L-PHENYLALANYL-TRNA(PHE).
CC   -!- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS
CC       (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; D64001; BAA10328.1; -.
DR   HSSP; P27001; 1PYS.
DR   PROSITE; PS00179; AA_TRNA_LIGASE_II_1; 1.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; 1.
DR   PFAM; PF01409; tRNA-synt_2d; 1.
KW   Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
SQ   SEQUENCE   331 AA;  37124 MW;  FA9E88E6 CRC32;

  Query Match             6.2%;  Score 99;  DB 1;  Length 331;
  Best Local Similarity  27.8%;  Pred. No. 1.42e+00;
  Matches    20;  Conservative   23;  Mismatches  24;  Indels   5;  Gaps   4;

Db     33 RVQYLGKKGELSLILKGMGKLSAE-ERPKFGAIANEVKEALQHDLESRKANLQNAAIEAQ 91
          || :||  ::|:  : :|::| :: :|: |  : : |  :|| : : | : :|  :: |
Qy     21 RVSFLG--AQLPPEVAAMARLLGDLDRSTFRKLLKFVVSSLQGE-DCRDG-VQRLGVSAN 76

Db     92 LAAETLDVTMAG 103
            |: | |   :||
Qy     77 LPEEQLGALLAG 88


RESULT   12
ID   FLIF_RHOSH     STANDARD;      PRT;   570 AA.
AC   Q53151;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-DEC-1999 (Rel. 39, Last annotation update)
DE   FLAGELLAR M-RING PROTEIN.
GN   FLIF.
OS   Rhodobacter sphaeroides (Rhodopseudomonas sphaeroides).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC   Rhodobacter.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WS8;
RA   GOODFELLOW I.G., WOOLLEY K.J., SOCKETT R.E.S.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE M RING MAY BE ACTIVELY INVOLVED IN ENERGY
CC       TRANSDUCTION (BY SIMILARITY).
```

```
CC   -!- SUBUNIT: THE BASAL BODY CONSTITUTES A MAJOR PORTION OF THE
CC       FLAGELLAR ORGANELLE AND CONSISTS OF FOUR RINGS (L,P,S, AND M)
CC       MOUNTED ON A CENTRAL ROD. THE M RING IS INTEGRAL TO THE INNER
CC       MEMBRANE OF THE CELL AND MAY BE CONNECTED TO THE FLAGELLAR ROD
CC       VIA THE S RING. THE S (SUPRAMEMBRANE RING) LIES JUST DISTAL TO
CC       THE M RING. THE L AND P RINGS LIE IN THE OUTER MEMBRANE AND THE
CC       PERIPLASMIC SPACE, RESPECTIVELY (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC, MEMBRANE-ASSOCIATED (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE FLIF FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X98692; CAA67251.1; -.
DR   PFAM; PF01514; YscJ_FliF; 1.
KW   Flagella; Membrane.
SQ   SEQUENCE   570 AA;  60675 MW;  1F8B8CE2 CRC32;

  Query Match           6.2%;  Score 99;  DB 1;  Length 570;
  Best Local Similarity  28.0%;  Pred. No. 1.42e+00;
  Matches   14;  Conservative   18;  Mismatches  17;  Indels   1;  Gaps   1;

Db    402 PPPPLLPEALKADLERLTRSAVGFDADRGDVVTITAQPFLDTVVPEASGW 451
          || :| |:::: :|: |  :   : :| : ||  : :|:||:|  : ::|
Qy    100 PPTSLKPDTFRDQLQELCIPQ-DLVGDLASVVFGSQRPLLDSVAQQQGAW 148


RESULT   13
ID   GCF_HUMAN      STANDARD;      PRT;   784 AA.
AC   P16383;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   01-JUN-1994 (Rel. 29, Last annotation update)
DE   GC-RICH SEQUENCE DNA-BINDING FACTOR (GCF) (TRANSCRIPTION FACTOR 9)
DE   (TCF-9).
GN   TCF9.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90075226.
RA   KAGEYAMA R., PASTAN I.;
RT   "Molecular cloning and characterization of a human DNA binding factor
RT   that represses transcription.";
RL   Cell 59:815-825(1989).
CC   -!- FUNCTION: FACTOR THAT REPRESSES TRANSCRIPTION. IT BINDS TO THE
CC       GC-RICH SEQUENCES (GCGGGGC) PRESENT IN THE EPIDERMAL GROWTH FACTOR
CC       RECEPTOR, BETA-ACTIN, AND CALCIUM-DEPENDENT PROTEASE PROMOTERS.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- TISSUE SPECIFICITY: WIDELY EXPRESSED IN HUMAN TISSUES AND CELL
CC       LINES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M29204; AAA35598.1; -.
DR   PIR; A33633; A33633.
DR   TRANSFAC; T00320; -.
DR   MIM; 189901; -.
KW   Transcription regulation; Repressor; DNA-binding; Trans-acting factor;
KW   Nuclear protein.
FT   DNA_BIND      1     78
FT   DOMAIN        1     86       ARG/LYS-RICH (BASIC).
FT   DOMAIN       23     29       POLY-LYS.
FT   DOMAIN      186    230       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN      359    380       LEUCINE-ZIPPER.
FT   DOMAIN      719    740       LEUCINE-ZIPPER.
SQ   SEQUENCE   784 AA;  91001 MW;  FE96504B CRC32;

  Query Match           6.2%;  Score 100;  DB 1;  Length 784;
  Best Local Similarity  31.4%;  Pred. No. 1.07e+00;
  Matches   16;  Conservative   12;  Mismatches  21;  Indels   2;  Gaps   2;

Db    485 PDSYYEAFISLCIPK-LLNPLIRVQLIDWNPLKLESTGLKEMPWFKSVEEF 534
          ||:: : :  ||||  |:  |  | :    || |:| :  : :|:  | :|
Qy    106 PDTFRDQLQELCIPQDLVGDLASVVFGSQRPL-LDSVAQQQGAWLPHVADF 155


RESULT   14
ID   YIN7_YEAST     STANDARD;      PRT;   946 AA.
AC   P40462;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   HYPOTHETICAL ZINC AMINOPEPTIDASE YIL137C (EC 3.4.11.-).
GN   YIL137C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   BARRELL B.G., BADCOCK K., BANKIER A.T., BOWMAN S., BROWN D.,
RA   CHURCHER C.M., CONNOR R., COPSEY T., DEAR S., DEVLIN K., FRASER A.,
RA   GENTLES S., HAMLYN N., HORSNELL T.S., HUNT S., JAGELS K., JONES M.,
RA   LOUIS E., LYE G., MOULE S., MOULE T., ODELL C., PEARSON D.,
RA   RAJANDREAM M.A., RILES L., ROWLEY N., SKELTON J., SMITH V.,
RA   WALSH S.V., WHITEHEAD S.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M1 (ZINC METALLOPROTEASE);
CC       ALSO KNOWN AS THE PEPN SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z38059; CAA86141.1; -.
DR   PIR; S48397; S48397.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
DR   PFAM; PF01433; Peptidase_M1; 1.
KW   Hypothetical protein; Hydrolase; Metalloprotease; Zinc;
KW   Aminopeptidase.
FT   METAL       330    330       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    331    331       BY SIMILARITY.
FT   METAL       334    334       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL       353    353       ZINC (CATALYTIC) (BY SIMILARITY).
SQ   SEQUENCE   946 AA;  107722 MW;  568EAFB6 CRC32;

  Query Match           6.2%;  Score 100;  DB 1;  Length 946;
  Best Local Similarity  31.0%;  Pred. No. 1.07e+00;
  Matches   22;  Conservative   21;  Mismatches  20;  Indels   8;  Gaps   8;

Db    299 GMITIQLNHLL-IPPNALANETVREQAQQL-IVHELVHQWMGNYISFDS-WESL-WFNES 354
          || |: | : | :|| :|  :| |:| |:| | ::|| : ::: : | | :  |    :
Qy     88 GMHTL-LQQALRLPPTSLKPDTFRDQLQELCIPQDLVGD-LAS-VVFGSQRPLLDSVAQQ 144

Db    355 FATWLACHILE 365
          ::||: |: :
Qy    145 QGAWLP-HVAD 154
```

```
RESULT   15
ID   RIR1_HSV23     STANDARD;      PRT;  1144 AA.
AC   P09853;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE LARGE CHAIN (EC 1.17.4.1)
DE   (RIBONUCLEOTIDE REDUCTASE) (136 KD SUBUNIT).
OS   Herpes simplex virus (type 2 / strain 333).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Simplexvirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 86144055.
RA   SWAIN M.A., GALLOWAY D.A.;
RT   "Herpes simplex virus specifies two subunits of ribonucleotide
RT   reductase encoded by 3'-coterminal transcripts.";
RL   J. Virol. 57:802-808(1986).
RN   [2]
RP   SEQUENCE OF 981-1144 FROM N.A.
RX   MEDLINE; 84138764.
RA   GALLOWAY D.A., SWAIN M.A.;
RT   "Organization of the left-hand end of the herpes simplex virus type 2
RT   BglII N fragment.";
RL   J. Virol. 49:724-730(1984).
RN   [3]
RP   SEQUENCE OF 1036-1144 FROM N.A.
RX   MEDLINE; 84057718.
RA   MCLAUCHLAN J., CLEMENTS J.B.;
RT   "DNA sequence homology between two co-linear loci on the HSV genome
RT   which have different transforming abilities.";
RL   EMBO J. 2:1953-1961(1983).
CC   -!- FUNCTION: PROVIDES THE PRECURSORS NECESSARY FOR DNA SYNTHESIS.
CC   -!- CATALYTIC ACTIVITY: 2'-DEOXYRIBONUCLEOSIDE DIPHOSPHATE + OXIDIZED
CC       THIOREDOXIN + H(2)O = RIBONUCLEOSIDE DIPHOSPHATE + REDUCED
CC       THIOREDOXIN.
CC   -!- PATHWAY: FIRST REACTION IN THE DNA REPLICATION PATHWAY.
CC   -!- SUBUNIT: HETERODIMER OF A LARGE AND A SMALL CHAIN.
CC   -!- SIMILARITY: BELONGS TO THE RIBONUCLEOSIDE DIPHOSPHATE REDUCTASE
CC       LARGE CHAIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M12700; AAA45806.1; -.
DR   EMBL; X00048; CAA24929.1; -.
DR   PIR; A05247; A05247.
DR   PROSITE; PS00089; RIBORED_LARGE; 1.
DR   PFAM; PF00317; ribonucleo_red; 1.
KW   Oxidoreductase; DNA replication; Early protein.
FT   DOMAIN      151    159       POLY-PRO.
FT   DOMAIN      224    231       POLY-ASP.
SQ   SEQUENCE   1144 AA;  125093 MW;  53E7FDD0 CRC32;

  Query Match             6.2%;  Score 100;  DB 1;  Length 1144;
  Best Local Similarity  29.7%;  Pred. No. 1.07e+00;
  Matches    27;  Conservative   24;  Mismatches  34;  Indels   6;  Gaps   5;

Db    324 ELTPENAEAVARFLGDAVDREP-ALMLEYFCRCAREESKRVPPRTFGSAPRLTEDDFGLL 382
          :| || | |:||:||| :||     :| :     : |  |   : :| :: | |:::| |
Qy     28 QLPPEVA-AMARLLGD-LDRSTFRKLLKFVVSSLQGEDCRDGVQRLGVSANLPEEQLGAL 85

Db    383 NYALAEMRRLCLDLPPVP--PNAYTPYHLRE 411
            ::  : :  | ||| :  |:::   :|:|
Qy     86 LAGMHTLLQQALRLPPTSLKPDTFRD-QLQE 115
```

```
Search completed: Fri May 12 10:32:10 2000
Job time : 40 secs.
```

****

MPSRCH (TM)

****

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 12 10:34:10 2000; MasPar time 54.28 Seconds
194.680 Million cell updates/sec
Tabular output not generated.

Title: >US-09-223-796-4
Description: (1-224) from US09223796.pep
Perfect Score: 1604
Sequence: 1 MSAVGAATPYLHHPGDSHSG.....ALVLKEMADLEKRCERRLQD 224

Scoring table: PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir62
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 46.697; Variance 102.828; scale 0.454

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	117	7.3	676	2	S41022	hypothetical protein	3.05e-02
2	110	6.9	2414	2	A54277	transcription adaptor	2.18e-01
3	109	6.8	234	2	S35020	nolW protein - Rhizob	2.87e-01
4	108	6.7	230	2	D64080	uroporphyrin-III C-me	3.78e-01
5	108	6.7	283	2	S25769	lipH protein - Pseudo	3.78e-01
6	107	6.7	451	2	S74728	hypothetical protein	4.96e-01
7	107	6.7	463	2	S77558	hypothetical protein	4.96e-01
8	106	6.6	339	2	S24161	lipase B - Pseudomona	6.50e-01
9	102	6.4	400	2	E70805	probable mce4 protein	1.88e+00
10	102	6.4	1116	2	S77213	DNA-directed DNA poly	1.88e+00
11	101	6.3	81	2	C69186	conserved hypothetica	2.45e+00
12	99	6.2	331	2	S74410	phenylalanine--tRNA l	4.11e+00
13	99	6.2	631	2	T15370	hypothetical protein	4.11e+00
14	100	6.2	784	2	A33633	transcription repress	3.18e+00
15	100	6.2	946	2	S48397	hypothetical protein	3.18e+00
16	98	6.1	368	2	S75923	sensory transduction	5.31e+00
17	98	6.1	434	2	S04534	invasin precursor - Y	5.31e+00
18	97	6.0	201	2	S11799	pullulanase secretion	6.85e+00
19	97	6.0	326	1	S11921	pulK protein - Klebsi	6.85e+00
20	97	6.0	479	2	B69764	transcription regulat	6.85e+00
21	97	6.0	482	1	A34671	triacylglycerol lipas	6.85e+00
22	97	6.0	661	2	A55877	tumor necrosis factor	6.85e+00
23	97	6.0	709	2	S75212	comE protein - Synech	6.85e+00
24	97	6.0	1111	2	T01239	hypothetical protein	6.85e+00
25	95	5.9	144	2	C70937	hypothetical protein	1.13e+01
26	95	5.9	189	1	S74659	shikimate kinase (EC	1.13e+01
27	95	5.9	221	2	C70936	probable IS1558 trans	1.13e+01
28	94	5.9	306	2	G64820	probable oligopeptide	1.45e+01
29	95	5.9	333	2	F70678	probable transposase	1.13e+01
30	94	5.9	500	1	JC4157	cytochrome P450 2D, e	1.45e+01
31	95	5.9	520	2	E71416	hypothetical protein	1.13e+01
32	94	5.9	579	2	F70000	two-component sensor	1.45e+01
33	95	5.9	603	2	JC1282	insulin-like growth f	1.13e+01
34	94	5.9	1137	1	WMBEB1	ribonucleoside-diphos	1.45e+01
35	95	5.9	2672	2	A48126	translation activator	1.13e+01
36	93	5.8	467	2	B65020	hypothetical protein	1.86e+01
37	93	5.8	560	2	S50392	hypothetical protein	1.86e+01
38	93	5.8	589	2	S65952	[phosphorylase] phosp	1.86e+01
39	93	5.8	603	2	JC6128	insulin-like growth f	1.86e+01
40	93	5.8	886	2	G65036	hypothetical protein	1.86e+01
41	93	5.8	1292	2	T09229	galactose binding adh	1.86e+01
42	93	5.8	1708	2	F69189	protoporphyrin IX mag	1.86e+01
43	93	5.8	2285	2	T13715	kakapo gene protein i	1.86e+01
44	93	5.8	2396	2	T13714	kakapo gene protein -	1.86e+01
45	93	5.8	4151	2	T13734	groovin gene protein	1.86e+01

ALIGNMENTS

```
RESULT    1
ENTRY           S41022     #type complete
TITLE           hypothetical protein T07C4.9 - Caenorhabditis elegans
ORGANISM        #formal_name Caenorhabditis elegans
DATE            06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change
                  09-Sep-1997
ACCESSIONS      S41022
REFERENCE       S41014
   #authors     Berks, M.
   #submission  submitted to the EMBL Data Library, January 1994
   #accession   S41022
      ##status       preliminary
      ##molecule_type DNA
      ##residues     1-676 ##label BER
      ##cross-references EMBL:Z29443; NID:g1067051; PID:g443836
GENETICS
   #introns     69/3; 161/1; 208/3; 227/1; 357/1; 505/3
CLASSIFICATION  #superfamily annexin repeat homology
FEATURE
   373-444          #domain annexin repeat homology #label AX1\
   445-516          #domain annexin repeat homology #label AX6\
   528-600          #domain annexin repeat homology #label AX3\
   604-675          #domain annexin repeat homology #label AX4
SUMMARY         #length 676  #molecular-weight 74968  #checksum 2702

  Query Match            7.3%;  Score 117;  DB 2;  Length 676;
  Best Local Similarity  22.5%;  Pred. No. 3.05e-02;
  Matches   16;  Conservative   26;  Mismatches  25;  Indels   4;  Gaps   4;

Db    132 EKVALERIVEQKLAESAKNEKDIAWNVNIVVANSYVAKSLRPVINIQMPTVGGDTNFEFD 191
          ::  |: : :|  :    :  |: | |:: :: | :|:||:| : :|:  ::  : : |:
Qy    134 QRPLLDSVAQQQ-GAWLPHVADFRWRVDVAISTSALARSLQPSVLMQL-KLSDGSAYRFE 191

Db    192 IDSFAQFRQQL 202
          : : | | |:|
Qy    192 VPT-AKF-QEL 200


RESULT    2
ENTRY           A54277     #type complete
TITLE           transcription adaptor protein p300 - human
ORGANISM        #formal_name Homo sapiens #common_name man
DATE            09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change
                  20-Sep-1999
ACCESSIONS      A54277
REFERENCE       A54277
   #authors     Eckner, R.; Ewen, M.E.; Newsome, D.; Gerdes, M.; DeCaprio,
```

```
                  J.A.; Lawrence, J.B.; Livingston, D.M.
   #journal       Genes Dev. (1994) 8:869-884
   #title         Molecular cloning and functional analysis of the adenovirus
                    E1A-associated 300-kD protein (p300) reveals a protein with
                    properties of a transcriptional adaptor.
   #cross-references MUID:95011587
   #accession     A54277
      ##status         preliminary
      ##molecule_type mRNA
      ##residues       1-2414 ##label ECK
      ##cross-references GB:U01877; NID:g495300; PIDN:AAA18639.1; PID:g495301
      ##note           in the authors' translation 941-Ser is shown after 961
                       and consequently, residues 942-961 are displaced one
                       codon to the left
GENETICS
   #map_position 22q13
CLASSIFICATION  #superfamily unassigned bromodomain proteins; bromodomain
                  homology
KEYWORDS        heptad repeat; zinc finger
FEATURE
   1075-1132          #domain bromodomain homology #label BRO
SUMMARY         #length 2414  #molecular-weight 264143  #checksum 9269

  Query Match             6.9%;  Score 110;  DB 2;  Length 2414;
  Best Local Similarity  34.0%;  Pred. No. 2.18e-01;
  Matches    17;  Conservative   13;  Mismatches  18;  Indels   2;  Gaps   2;

Db   2137 QAGVQRAGLPQQQPQQQLQPPMGGMSPQAQQ-MNMNHNTM-PSQFRDILR 2184
         : |||| |::   |::|| : ::||    || : :   :: |  ||| |:
Qy     65 RDGVQRLGVSANLPEEQLGALLAGMHTLLQQALRLPPTSLKPDTFRDQLQ 114


RESULT    3
ENTRY           S35020     #type complete
TITLE           nolW protein - Rhizobium fredii
ORGANISM        #formal_name Rhizobium fredii
DATE            03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change
                  09-Sep-1997
ACCESSIONS      S35020
REFERENCE       S35019
   #authors     Meinhardt, L.W.; Krishnan, H.B.; Balatti, P.A.; Pueppke, S.G.
   #journal     Mol. Microbiol. (1993) 9:17-29
   #title       Molecular cloning and characterization of a sym plasmid locus
                  that regulates cultivar-specific nodulation of soybean by
                  Rhizobium fredii USDA257.
   #cross-references MUID:94018604
   #accession   S35020
      ##molecule_type DNA
      ##residues      1-234 ##label MEI
      ##cross-references EMBL:L12251; NID:g1648932; PID:g152383
GENETICS
   #gene        nolW
KEYWORDS        membrane protein
SUMMARY         #length 234  #molecular-weight 25829  #checksum 725

  Query Match             6.8%;  Score 109;  DB 2;  Length 234;
  Best Local Similarity  27.0%;  Pred. No. 2.87e-01;
  Matches    34;  Conservative   35;  Mismatches  49;  Indels   8;  Gaps   8;

Db     24 LFAGIHTTLGATLPLPSTSYKYTVLDQDLSAALQEFGNNLKISVNISAEVKGRIRGRIPE 83
          |:||:|| |  :| ||:|| |   : ::|   |  :  :| :: :::: | |  |  :
Qy     85 LLAGMHTLLQQALRLPPTSLKPDTFRDQL-QELC-IPQDL-VG-DLASVVFGSQRPLLDS 140

Db     84 LSPREFLDRLTDLYDLQWYYDGVVLYVSA-AKEAQTRMLV-LSSVHFSAFKLALDKLDIS 141
          :: ::   :|  : |::|  | | :  || |:  |  :|: |     ||::: :    :
Qy    141 VAQQQG-AWLPHVADFRWRVD-VAISTSALARSLQPSVLMQLKLSDGSAYRFEVPTAKFQ 198

Db    142 DERYPV 147
          : ||:|
Qy    199 ELRYSV 204


RESULT    4
ENTRY           D64080     #type complete
TITLE           uroporphyrin-III C-methyltransferase homolog - Haemophilus
                  influenzae (strain Rd KW20)
ORGANISM        #formal_name Haemophilus influenzae
DATE            18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change
                  31-Oct-1997
ACCESSIONS      D64080
REFERENCE       A64000
   #authors     Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.;
                  Kirkness, E.F.; Kerlavage, A.R.; Bult, C.J.; Tomb, J.F.;
                  Dougherty, B.A.; Merrick, J.M.; McKenney, K.; Sutton, G.;
                  FitzHugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.;
                  Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman,
                  J.F.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton,
                  M.D.; Utterback, T.R.; Hanna, M.C.; Nguyen, D.T.; Saudek,
                  D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann,
                  J.L.; Geoghagen, N.S.M.; Gnehm, C.L.; McDonald, L.A.;
                  Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.C.
   #journal     Science (1995) 269:496-512
   #title       Whole-genome random sequencing and assembly of Haemophilus
                  influenzae Rd.
   #cross-references MUID:95350630
   #accession   D64080
      ##status       nucleic acid sequence not shown; translation not shown
      ##molecule_type DNA
      ##residues      1-230 ##label TIGR
      ##cross-references GB:U32742; GB:L42023; NID:g1573593; PID:g1573595;
                       TIGR:HI0603
SUMMARY         #length 230  #molecular-weight 25193  #checksum 4558

  Query Match             6.7%;  Score 108;  DB 2;  Length 230;
  Best Local Similarity  27.5%;  Pred. No. 3.78e-01;
  Matches    38;  Conservative   34;  Mismatches  56;  Indels  10;  Gaps   7;

Db     76 ENQTGANLSSNNTNNNKRLTQLEQSLKTAQENIAQLEQLIVSK--TGEITSLQTQMKQVS 133
          |:| || |:: :|     |     ||| ::    ||::| ::    |:::|:     | :
Qy     79 EEQLGALLAGMHTLLQQALRLPPTSLK-PDTFRDQLQELCIPQDLVGDLASVVFG-SQRP 136

Db    134 QL-AIAQQPSDWLFSEADFL--LNNALRKLVLDNDVDTAVSL-LKLADETLVK--VNNSQ 187
           | ::||| : ||   |||   :: |:    |   :: :| : |||:| :  :  |  :
Qy    137 LLDSVAQQQGAWLPHVADFRWRVDVAISTSALARSLQPSVLMQLKLSDGSAYRFEVPTAK 196

Db    188 ANEIRSAINQDLKQLLSL 205
            |:| ::   ||::  |
Qy    197 FQELRYSVALVLKEMADL 214


RESULT    5
ENTRY           S25769     #type complete
TITLE           lipH protein - Pseudomonas aeruginosa
ORGANISM        #formal_name Pseudomonas aeruginosa
DATE            13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
                  07-May-1999
ACCESSIONS      S25769
REFERENCE       S25768
   #authors     Wohlfarth, S.; Hoesche, C.; Strunk, C.; Winkler, U.K.
   #journal     J. Gen. Microbiol. (1992) 138:1325-1335
   #title       Molecular genetics of the extracellular lipase of Pseudomonas
                  aeruginosa PAO1.
   #cross-references MUID:92381478
   #accession   S25769
      ##status       preliminary
      ##molecule_type DNA
      ##residues      1-283 ##label WOH
      ##cross-references EMBL:X63391
SUMMARY         #length 283  #molecular-weight 31853  #checksum 5389

  Query Match             6.7%;  Score 108;  DB 2;  Length 283;
  Best Local Similarity  30.5%;  Pred. No. 3.78e-01;
  Matches    18;  Conservative   22;  Mismatches  16;  Indels   3;  Gaps   2;

Db    150 LSAEEKAAAIDRL--RASLPEDQQESVLPQLQSELQQQTAALQAAGAGPEAIRQMRQQL 206
          | :|:   :::||   |:|||:|  ::|: ::: | ||:  | :::  |:::|:  |:|
```

```
Qy      59 LQGEDCRDGVQRLGVSANLPEEQLGALLAGMHT-LLQQALRLPPTSLKPDTFRDQLQEL 116


RESULT    6
ENTRY             S74728     #type complete
TITLE             hypothetical protein slr0971 ccmA 3'-region - Synechocystis
                    sp. (strain PCC 6803)
ORGANISM          #formal_name Synechocystis sp.
   #variety       PCC 6803
DATE              25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
                    21-Aug-1998
ACCESSIONS        S74728; B55214
REFERENCE         S74322
   #authors       Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
                    Nakamura, Y.; Miyajima, N.; Hirosawa, M.; Sugiura, M.;
                    Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
                    Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo,
                    S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
                    Yasuda, M.; Tabata, S.
   #journal       DNA Res. (1996) 3:109-136
   #title         Sequence analysis of the genome of the unicellular
                    cyanobacterium Synechocystis sp. PCC6803. II. Sequence
                    determination of the entire genome and assignment of
                    potential protein-coding regions.
   #cross-references MUID:97061201
   #accession     S74728
      ##status         nucleic acid sequence not shown; translation not shown
      ##molecule_type DNA
      ##residues       1-451 ##label KAN
      ##cross-references EMBL:D90901; GB:AB001339; NID:g1651897; PID:d1017612;
                          PID:g1651953
      ##note           the nucleotide sequence was submitted to the EMBL Data
                        Library, June 1996
REFERENCE         A55214
   #authors       Ogawa, T.; Marco, E.; Orus, M.I.
   #journal       J. Bacteriol. (1994) 176:2374-2378
   #title         A gene (ccmA) required for carboxysome formation in the
                    cyanobacterium Synechocystis sp. strain PCC6803.
   #cross-references MUID:94209239
   #accession     B55214
      ##molecule_type DNA
      ##residues       'VDPNLEVAQISNRMLQAGGPGLLFENVKG',
                        'SPFPVAVNLMGTVERICWAMNMDHPLELEDLGKK',244-451 ##label
                        OGA
      ##cross-references GB:D26444; NID:g452237; PID:d1006006; PID:g499299
      ##note           the authors translated the codon CCC for residue 236 as
                        Glu; the authors translated used GTG as an initiation
                        codon and translated it as Val
      ##note           the authors say that codon usages suggest this protein
                        is not expressed at high levels
SUMMARY           #length 451  #molecular-weight 49667  #checksum 7763

  Query Match             6.7%;  Score 107;  DB 2;  Length 451;
  Best Local Similarity  38.6%;  Pred. No. 4.96e-01;
  Matches    22;  Conservative   13;  Mismatches  18;  Indels   4;  Gaps   4;

Db      42 ALHFDDPRLPEVEQLTLSGT-PEE-LEALALAVETYLQQRLN-PCASLEPSTYPEPV 95
           :|: :| |   |::| :|:  ||| | ||  :: | ||| |  | :|| | |: : :
Qy      58 SLQGEDCR-DGVQRLGVSANLPEEQLGALLAGMHTLLQQALRLPPTSLKPDTFRDQL 113


RESULT    7
ENTRY             S77558     #type complete
TITLE             hypothetical protein sll1967 - Synechocystis sp. (strain PCC
                    6803)
ORGANISM          #formal_name Synechocystis sp.
   #variety       PCC 6803
DATE              25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
                    26-Aug-1999
ACCESSIONS        S77558
REFERENCE         S74322
   #authors       Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
                    Nakamura, Y.; Miyajima, N.; Hirosawa, M.; Sugiura, M.;
                    Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
                    Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo,
                    S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
                    Yasuda, M.; Tabata, S.
   #journal       DNA Res. (1996) 3:109-136
   #title         Sequence analysis of the genome of the unicellular
                    cyanobacterium Synechocystis sp. PCC6803. II. Sequence
                    determination of the entire genome and assignment of
                    potential protein-coding regions.
   #cross-references MUID:97061201
   #accession     S77558
      ##status         nucleic acid sequence not shown; translation not shown
      ##molecule_type DNA
      ##residues       1-463 ##label KAN
      ##cross-references EMBL:D90905; GB:AB001339; NID:g1652360; PID:d1018138;
                          PID:g1652483
      ##note           the nucleotide sequence was submitted to the EMBL Data
                        Library, June 1996
CLASSIFICATION    #superfamily hypothetical protein HI0333
SUMMARY           #length 463  #molecular-weight 51523  #checksum 8580

  Query Match             6.7%;  Score 107;  DB 2;  Length 463;
  Best Local Similarity  22.3%;  Pred. No. 4.96e-01;
  Matches    31;  Conservative   41;  Mismatches  57;  Indels  10;  Gaps  10;

Db     116 RIGGFTTLPLEPLLQSPASL-GYRNKATYP-LSRSKTGQVQAGYYRKGSHRLVNINQCPV 173
           |:: |    | | : : | | |  :::|:  | :   : :|:   | | :|| ::   :
Qy      21 RVS-FLGAQLPPEVAAMARLLGDLDRSTFRKLLKFVVSSLQGEDCRDGVQRL-GVSA-NL 77

Db     174 QDDRLNLLLTEVKKDIENRGWSIYDEEKKQGKLR-HL-SLRIGQR-TGEMLLTLISAHKG 230
            :::|  ||: :  :: :: :    |  :| :|  | | |   |::  ::::::
Qy      78 PEEQLGALLAGMHTLLQ-QALRLPPTSLKPDTFRDQLQELCIPQDLVGDLASVVFGSQRP 136

Db     231 LPD-LEEQAGEWLERYPDL 248
           | | : :| | || : :|:
Qy     137 LLDSVAQQQGAWLPHVADF 155


RESULT    8
ENTRY             S24161     #type complete
TITLE             lipase B - Pseudomonas aeruginosa
ORGANISM          #formal_name Pseudomonas aeruginosa
DATE              19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change
                    20-Mar-1998
ACCESSIONS        S24161
REFERENCE         S24160
   #authors       Chihara-Siomi, M.; Yoshikawa, K.; Oshima-Hirayama, N.;
                    Yamamoto, K.; Sogabe, Y.; Nakatani, T.; Nishioka, T.; Oda,
                    J.
   #journal       Arch. Biochem. Biophys. (1992) 296:505-513
   #title         Purification, molecular cloning, and expression of lipase
                    from Pseudomonas aeruginosa.
   #cross-references MUID:92337414
   #accession     S24161
      ##status         preliminary
      ##molecule_type DNA
      ##residues       1-339 ##label CHI
      ##cross-references GB:D10048; NID:g216895; PID:d1001403; PID:g216897
SUMMARY           #length 339  #molecular-weight 37532  #checksum 1707

  Query Match             6.6%;  Score 106;  DB 2;  Length 339;
  Best Local Similarity  30.5%;  Pred. No. 6.50e-01;
  Matches    18;  Conservative   21;  Mismatches  17;  Indels   3;  Gaps   2;

Db     207 LSTEEKAAAIDRL--RASLPEDQQESVLPQLQSELQQQTAALQAAGAGPEAIRQMRQQL 263
           |  |:   :::||   |:|||:|  ::|: ::: | ||:  | ::: |:::|:  |:|
Qy      59 LQGEDCRDGVQRLGVSANLPEEQLGALLAGMHT-LLQQALRLPPTSLKPDTFRDQLQEL 116


RESULT    9
ENTRY             E70805     #type complete
TITLE             probable mce4 protein - Mycobacterium tuberculosis (strain
                    H37RV)
```

```
ORGANISM          #formal_name Mycobacterium tuberculosis
DATE              17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
                    17-Jul-1998
ACCESSIONS        E70805
REFERENCE         A70500
   #authors       Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
                    C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry
                    III, C.E.; Tekaia, F.; Badcock, K.; Basham, D.; Brown, D.;
                    Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
                    Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
                    Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
                    Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
                    Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
                    Skelton, S.; Squares, S.; Sqares, R.; Sulston, J.E.;
                    Taylor, K.; Whitehead, S.; Barrell, B.G.
   #journal       Nature (1998) 393:537-544
   #title         Deciphering the biology of Mycobacterium tuberculosis from
                    the complete genome sequence.
   #cross-references MUID:98295987
   #accession     E70805
      ##status         preliminary; nucleic acid sequence not shown;
                         translation not shown
      ##molecule_type DNA
      ##residues       1-400 ##label COL
      ##cross-references GB:AL022022; GB:AL123456; NID:g3261554; PID:e1254627;
                          PID:g2924436
      ##experimental_source strain H37Rv
GENETICS
   #gene          mce4
SUMMARY           #length 400  #molecular-weight 42418  #checksum 124

  Query Match             6.4%;  Score 102;  DB 2;  Length 400;
  Best Local Similarity  26.3%;  Pred. No. 1.88e+00;
  Matches    31;  Conservative   31;  Mismatches  48;  Indels   8;  Gaps   7;

Db    126 LSPNAHVAASQV-QLEVNTLFQSLIDLL-HKIDPLETNATLSALSEGLRGHGDDLGALLS 183
          |:|:    |  : :|: :| |: |: ::   ::  :    :  |: :     ::||||||:
Qy     29 LPPEVAAMARLLGDLDRST-FRKLLKFVVSSLQGEDCRDGVQRLGVSANLPEEQLGALLA 87

Db    184 GLNTLTRQANPKLP--ALQED-FRKAAVVANVYADAAGDLNTV-FDNLPTINKTIVDQ 237
          |::|| :|| :||  :|  | ||       :  |  ||| :| | :   :  :: :|
Qy     88 GMHTLLQQA-LRLPPTSLKPDTFRDQLQELCIPQDLVGDLASVVFGSQRPLLDSVAQQ 144


RESULT   10
ENTRY             S77213     #type complete
TITLE             DNA-directed DNA polymerase (EC 2.7.7.7) III chain dnaX -
                    Synechocystis sp. (strain PCC 6803)
ALTERNATE_NAMES   protein sll1360
ORGANISM          #formal_name Synechocystis sp.
   #variety       PCC 6803
DATE              25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
                    21-Aug-1998
ACCESSIONS        S77213
REFERENCE         S74322
   #authors       Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
                    Nakamura, Y.; Miyajima, N.; Hirosawa, M.; Sugiura, M.;
                    Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
                    Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo,
                    S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
                    Yasuda, M.; Tabata, S.
   #journal       DNA Res. (1996) 3:109-136
   #title         Sequence analysis of the genome of the unicellular
                    cyanobacterium Synechocystis sp. PCC6803. II. Sequence
                    determination of the entire genome and assignment of
                    potential protein-coding regions.
   #cross-references MUID:97061201
   #accession     S77213
      ##status         nucleic acid sequence not shown; translation not shown
      ##molecule_type DNA
      ##residues       1-1116 ##label KAN
      ##cross-references EMBL:D90907; GB:AB001339; NID:g1652618; PID:d1018280;
                          PID:g1652627
      ##note           the nucleotide sequence was submitted to the EMBL Data
                        Library, June 1996
GENETICS
   #gene          dnaX
KEYWORDS          nucleotidyltransferase
SUMMARY           #length 1116  #molecular-weight 122009  #checksum 8010

  Query Match             6.4%;  Score 102;  DB 2;  Length 1116;
  Best Local Similarity  23.3%;  Pred. No. 1.88e+00;
  Matches    30;  Conservative   29;  Mismatches  66;  Indels   4;  Gaps   4;

Db    576 EEPPERVVFVLATTDPQ-RVLPTIISRCQRFDYRRIPLQAMVDHLRYIAGRENINIDQPA 634
          :  : || |: |   |:   :: :::  :|  :|:: |  :|  |:    |:|:    :
Qy     16 DSHSGRVSFLGAQLPPEVAAMARLLGDLDRSTFRKL-LKFVVSSLQGEDCRDGVQRLGVS 74

Db    635 LTLVAQIANGGLRDAESLLDQ-LSLLPDLITPDKVWD-LVGAVPEQDLLALLEAIASDDA 692
            |  :   : |    :||:| | | |  : ||  :| |      |||:: || ::
Qy     75 ANLPEEQLGALLAGMHTLLQQALRLPPTSLKPDTFRDQLQELCIPQDLVGDLASVVFGSQ 134

Db    693 ETLLATCRQ 701
            || :  |
Qy    135 RPLLDSVAQ 143


RESULT   11
ENTRY             C69186     #type complete
TITLE             conserved hypothetical protein MTH649 - Methanobacterium
                    thermoautotrophicum (strain Delta H)
ORGANISM          #formal_name Methanobacterium thermoautotrophicum
DATE              05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
                    18-Sep-1998
ACCESSIONS        C69186
REFERENCE         A69000
   #authors       Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.;
                    Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.;
                    Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.;
                    Lumm, W.; Pothier, B.; Qiu, D.; Spadafora, R.; Vicaire, R.;
                    Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Caruso,
                    A.; Bush, D.; Safer, H.; Patwell, D.; Prabhakar, S.;
                    McDougall, S.; Shimer, G.; Goyal, A.; Pietrokovski, S.;
                    Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling,
                    J.; Reeve, J.N.
   #journal       J. Bacteriol. (1997) 179:7135-7155
   #title         Complete genome sequence of Methanobacterium
                    thermoautotrophicum Delta H: functional analysis and
                    comparative genomics.
   #cross-references MUID:98037514
   #accession     C69186
      ##status         preliminary; nucleic acid sequence not shown;
                         translation not shown
      ##molecule_type DNA
      ##residues       1-81 ##label MTH
      ##cross-references GB:AE000845; GB:AE000666; NID:g2621726; PID:g2621731
      ##experimental_source strain Delta H
GENETICS
   #gene          MTH649
   #start_codon   GTG
CLASSIFICATION    #superfamily yeast SNP2 protein
SUMMARY           #length 81  #molecular-weight 9061  #checksum 8046

  Query Match             6.3%;  Score 101;  DB 2;  Length 81;
  Best Local Similarity  34.8%;  Pred. No. 2.45e+00;
  Matches    23;  Conservative   16;  Mismatches  19;  Indels   8;  Gaps   7;

Db      8 RVNVQRPLDALGNSLNSPVIIKLK-GD-REFR-GVLKS-F-DLH--MNLVLNDAEELEDG 60
          ||:|  :  ||: || ::|:: || :|   :|  |  : | :|:  : |||::  :||
Qy    158 RVDVAISTSALARSLQPSVLMQLKLSDGSAYRFEVPTAKFQELRYSVALVLKEMADLEK- 216

Db     61 EVTRRL 66
            |||
Qy    217 RCERRL 222
```

```
RESULT   12
ENTRY            S74410     #type complete
TITLE            phenylalanine--tRNA ligase (EC 6.1.1.20) alpha chain -
                   Synechocystis sp. (PCC 6803)
ALTERNATE_NAMES  phenylalanyl-tRNA synthetase alpha chain; protein sll0454
ORGANISM         #formal_name Synechocystis sp.
   #variety      PCC 6803
DATE             25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
                   16-Jul-1999
ACCESSIONS       S74410
REFERENCE        S74322
   #authors      Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
                   Nakamura, Y.; Miyajima, N.; Hirosawa, M.; Sugiura, M.;
                   Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
                   Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo,
                   S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
                   Yasuda, M.; Tabata, S.
   #journal      DNA Res. (1996) 3:109-136
   #title        Sequence analysis of the genome of the unicellular
                   cyanobacterium Synechocystis sp. PCC6803. II. Sequence
                   determination of the entire genome and assignment of
                   potential protein-coding regions.
   #cross-references MUID:97061201
   #accession    S74410
      ##status        nucleic acid sequence not shown; translation not shown
      ##molecule_type DNA
      ##residues      1-331 ##label KAN
      ##cross-references EMBL:D64001; GB:AB001339; NID:g1001102;
                        PIDN:BAA10328.1; PID:d1010979; PID:g1001185
      ##experimental_source PCC 6803
      ##note          the nucleotide sequence was submitted to the EMBL Data
                       Library, June 1996
GENETICS
   #gene         pheS
FUNCTION
   #description  catalyzes transfer of activated phenylalanine to
                   phenylalanyl-tRNA
CLASSIFICATION   #superfamily phenylalanine--tRNA ligase alpha chain
KEYWORDS         aminoacyl-tRNA synthetase; ATP; ligase; protein biosynthesis
SUMMARY          #length 331  #molecular-weight 37124  #checksum 6487

  Query Match             6.2%;  Score 99;  DB 2;  Length 331;
  Best Local Similarity  27.8%;  Pred. No. 4.11e+00;
  Matches    20;  Conservative   23;  Mismatches  24;  Indels   5;  Gaps   4;

Db      33 RVQYLGKKGELSLILKGMGKLSAE-ERPKFGAIANEVKEALQHDLESRKANLQNAAIEAQ 91
           || :||  ::|:  : :|::| :: :|: |  : : |  :|| : : | : :|  :: |
Qy      21 RVSFLG--AQLPPEVAAMARLLGDLDRSTFRKLLKFVVSSLQGE-DCRDG-VQRLGVSAN 76

Db      92 LAAETLDVTMAG 103
           |: | |   :||
Qy      77 LPEEQLGALLAG 88


RESULT   13
ENTRY            T15370     #type complete
TITLE            hypothetical protein C01F1.3 - Caenorhabditis elegans
ORGANISM         #formal_name Caenorhabditis elegans
DATE             20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change
                   20-Sep-1999
ACCESSIONS       T15370
REFERENCE        Z18338
   #authors      Johnson, D.
   #submission   submitted to the EMBL Data Library, May 1996
   #description  The sequence of C. elegans cosmid C01F1.
   #accession    T15370
      ##status        preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues      1-631 ##label JOH
      ##cross-references EMBL:U58761; NID:g1330391; PID:g1330395;
                        PIDN:AAB00715.1; GSPDB:GN00020; CESP:C01F1.3
      ##experimental_source strain Bristol N2; clone C01F1
GENETICS
   #gene         CESP:C01F1.3
   #map_position 2
   #introns      80/3; 375/3; 423/2; 464/1; 490/3
SUMMARY          #length 631  #molecular-weight 71113  #checksum 1765

  Query Match             6.2%;  Score 99;  DB 2;  Length 631;
  Best Local Similarity  26.9%;  Pred. No. 4.11e+00;
  Matches    14;  Conservative   17;  Mismatches  18;  Indels   3;  Gaps   3;

Db     258 NVADLAKTIQEEVDLQLGRAHEPPKYK-SIPDRPYNDLRYLIS-IEKAKNDL 307
           : : ||:::|  | :|| :  : : |:  :|   : :||| :: : | : ||
Qy     164 STSALARSLQPSVLMQL-KLSDGSAYRFEVPTAKFQELRYSVALVLKEMADL 214


RESULT   14
ENTRY            A33633     #type complete
TITLE            transcription repressor protein GCF - human
ORGANISM         #formal_name Homo sapiens #common_name man
DATE             02-Feb-1990 #sequence_revision 02-Feb-1990 #text_change
                   10-Sep-1997
ACCESSIONS       A33633
REFERENCE        A33633
   #authors      Kageyama, R.; Pastan, I.
   #journal      Cell (1989) 59:815-825
   #title        Molecular cloning and characterization of a human DNA binding
                   factor that represses transcription.
   #cross-references MUID:90075226
   #accession    A33633
      ##status        preliminary
      ##molecule_type mRNA
      ##residues      1-784 ##label KAG
      ##cross-references GB:M29204; NID:g179411; PID:g179412
KEYWORDS         DNA binding; transcription regulation
SUMMARY          #length 784  #molecular-weight 91001  #checksum 3428

  Query Match             6.2%;  Score 100;  DB 2;  Length 784;
  Best Local Similarity  31.4%;  Pred. No. 3.18e+00;
  Matches    16;  Conservative   12;  Mismatches  21;  Indels   2;  Gaps   2;

Db     485 PDSYYEAFISLCIPK-LLNPLIRVQLIDWNPLKLESTGLKEMPWFKSVEEF 534
           ||:: : :  ||||  |:  |  | :    || |:| :  : :|:  | :|
Qy     106 PDTFRDQLQELCIPQDLVGDLASVVFGSQRPL-LDSVAQQQGAWLPHVADF 155


RESULT   15
ENTRY            S48397     #type complete
TITLE            hypothetical protein YIL137c - yeast (Saccharomyces
                   cerevisiae)
ORGANISM         #formal_name Saccharomyces cerevisiae
DATE             02-Dec-1994 #sequence_revision 02-Dec-1994 #text_change
                   21-Nov-1997
ACCESSIONS       S48397
REFERENCE        S48310
   #authors      Churcher, C.
   #submission   submitted to the EMBL Data Library, September 1994
   #accession    S48397
      ##molecule_type DNA
      ##residues      1-946 ##label CHU
      ##cross-references GB:Z47047; EMBL:Z38059; NID:g603997; PID:g763209;
                        MIPS:YIL137c
GENETICS
   #map_position 9L
SUMMARY          #length 946  #molecular-weight 107722  #checksum 5110

  Query Match             6.2%;  Score 100;  DB 2;  Length 946;
  Best Local Similarity  31.0%;  Pred. No. 3.18e+00;
  Matches    22;  Conservative   21;  Mismatches  20;  Indels   8;  Gaps   8;

Db     299 GMITIQLNHLL-IPPNALANETVREQAQQL-IVHELVHQWMGNYISFDS-WESL-WFNES 354
           || |: | : | :|| :|  :| |:| |:| | ::|| : ::: : | | :  |    :
Qy      88 GMHTL-LQQALRLPPTSLKPDTFRDQLQELCIPQDLVGD-LAS-VVFGSQRPLLDSVAQQ 144

Db     355 FATWLACHILE 365
```

::||: |: :
Qy 145 QGAWLP-HVAD 154

Search completed: Fri May 12 10:35:10 2000
Job time : 60 secs.

****

SSEARCH (TM)

****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 12 10:36:18 2000; MasPar time 17.97 Seconds
161.524 Million cell updates/sec
Tabular output not generated.

Title: >US-09-223-796-4
Description: (1-224) from US09223796.pep
Perfect Score: 1604
Sequence: 1 MSAVGAATPYLHHPGDSHSG.....ALVLKEMADLEKRCERRLQD 224

Scoring table: PAM 150
Gap 11

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT_COMB 4:backfiles1

Statistics: Mean 30.751; Variance 151.832; scale 0.203

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	125	7.8	195	2	US-08-822-	Sequence 3, Applicatio	2.01e-01
2	122	7.6	195	2	US-08-822-	Sequence 1, Applicatio	3.27e-01
3	110	6.9	2414	3	PCT-US95-0	Sequence 2, Applicatio	2.22e+00
4	110	6.9	2414	1	US-08-227-	Sequence 2, Applicatio	2.22e+00
5	107	6.7	395	2	US-08-795-	Sequence 18, Applicati	3.56e+00
6	107	6.7	395	1	US-08-318-	Sequence 18, Applicati	3.56e+00
7	107	6.7	1114	2	US-08-576-	Sequence 31, Applicati	3.56e+00
8	100	6.2	448	2	US-09-015-	Sequence 1, Applicatio	1.05e+01
9	98	6.1	1064	2	US-09-003-	Sequence 2, Applicatio	1.43e+01
10	98	6.1	1064	1	US-08-357-	Sequence 2, Applicatio	1.43e+01
11	98	6.1	1064	3	PCT-US95-1	Sequence 2, Applicatio	1.43e+01
12	98	6.1	1082	2	US-09-003-	Sequence 5, Applicatio	1.43e+01
13	98	6.1	1082	1	US-08-357-	Sequence 10, Applicati	1.43e+01
14	98	6.1	1082	3	PCT-US95-1	Sequence 5, Applicatio	1.43e+01
15	98	6.1	1082	3	PCT-US95-1	Sequence 10, Applicati	1.43e+01
16	98	6.1	1082	1	US-08-357-	Sequence 5, Applicatio	1.43e+01
17	98	6.1	1082	2	US-09-003-	Sequence 10, Applicati	1.43e+01
18	94	5.9	382	2	US-08-795-	Sequence 17, Applicati	2.61e+01
19	94	5.9	382	1	US-08-318-	Sequence 17, Applicati	2.61e+01
20	95	5.9	603	1	US-08-190-	Sequence 50, Applicati	2.25e+01
21	93	5.8	155	2	US-08-576-	Sequence 55, Applicati	3.04e+01
22	92	5.7	335	2	US-08-875-	Sequence 3, Applicatio	3.53e+01
23	92	5.7	350	3	PCT-US95-0	Sequence 14, Applicati	3.53e+01
24	92	5.7	350	2	US-08-765-	Sequence 14, Applicati	3.53e+01
25	91	5.7	476	1	US-08-565-	Sequence 6, Applicatio	4.09e+01
26	91	5.7	490	2	US-08-946-	Sequence 2, Applicatio	4.09e+01
27	91	5.7	490	1	US-08-361-	Sequence 2, Applicatio	4.09e+01
28	91	5.7	490	1	US-08-565-	Sequence 2, Applicatio	4.09e+01
29	91	5.7	3031	1	US-07-689-	Sequence 2, Applicatio	4.09e+01
30	88	5.5	454	2	US-09-014-	Sequence 9, Applicatio	6.38e+01
31	89	5.5	494	1	US-08-447-	Sequence 2, Applicatio	5.51e+01
32	89	5.5	494	1	US-08-453-	Sequence 2, Applicatio	5.51e+01
33	89	5.5	494	1	US-08-447-	Sequence 2, Applicatio	5.51e+01
34	89	5.5	494	1	US-08-454-	Sequence 2, Applicatio	5.51e+01
35	87	5.4	419	1	US-08-385-	Sequence 2, Applicatio	7.39e+01
36	86	5.4	560	1	US-07-683-	Sequence 1, Applicatio	8.55e+01
37	86	5.4	963	1	US-08-537-	Sequence 3, Applicatio	8.55e+01
38	86	5.4	1143	2	US-08-310-	Sequence 108, Applicat	8.55e+01
39	86	5.4	1143	3	PCT-US95-0	Sequence 108, Applicat	8.55e+01
40	86	5.4	1144	1	US-08-261-	Sequence 2, Applicatio	8.55e+01
41	86	5.4	1144	3	PCT-US95-0	Sequence 4, Applicatio	8.55e+01
42	86	5.4	1144	1	US-08-261-	Sequence 4, Applicatio	8.55e+01
43	86	5.4	1144	3	PCT-US95-0	Sequence 2, Applicatio	8.55e+01
44	84	5.2	394	1	US-08-002-	Sequence 9, Applicatio	1.14e+02
45	84	5.2	394	1	US-08-002-	Sequence 6, Applicatio	1.14e+02

ALIGNMENTS

```
RESULT    1
ID   US-08-822-260-3     STANDARD;     PRT;   195 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 3, Application US/08822260
XX
CC    Sequence 3, Application US/08822260
CC    Patent No. 5830660
CC     GENERAL INFORMATION:
CC       APPLICANT:  Hillman, Jennifer L.
CC       APPLICANT:  Goli, Surya K.
CC       TITLE OF INVENTION:  NOVEL TUMORIGENESIS PROTEIN
CC       NUMBER OF SEQUENCES:  3
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Incyte Pharmaceuticals, Inc.
CC         STREET:  3174 Porter Drive
CC         CITY:  Palo Alto
CC         STATE:  CA
CC         COUNTRY:  USA
CC         ZIP:  94304
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette
CC         COMPUTER:  IBM Compatible
CC         OPERATING SYSTEM:  DOS
CC         SOFTWARE:  FastSEQ for Windows Version 2.0
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/822,260
CC         FILING DATE:  Herewith
CC         CLASSIFICATION:  514
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:
CC         FILING DATE:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Billings, Lucy J.
CC         REGISTRATION NUMBER:  36,749
CC         REFERENCE/DOCKET NUMBER:  PF-0247 US
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  415-855-0555
CC         TELEFAX:  415-845-4166
CC         TELEX:
CC     INFORMATION FOR SEQ ID NO:  3:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  195 amino acids
CC         TYPE:  amino acid
```

```
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       IMMEDIATE SOURCE:
CC         LIBRARY:  GenBank
CC         CLONE:  265569
SQ   SEQUENCE   195 AA; 22037 MW; 190094 CN;

  Query Match            7.8%;  Score 125;  DB 2;  Length 195;
  Best Local Similarity  22.2%;  Pred. No. 2.01e-01;
  Matches    24;  Conservative   37;  Mismatches  40;  Indels   7;  Gaps   7;

Db     87 LED-CKFDRERI-ELFCTEYQNNKNSLETLLGSIGRSLPHITDVSWRLEYQIKTNQLHKM 144
          |:: | : :: : :|    : : :  |:::    |  |||::|  ||::  | |: | :
Qy    113 LQELC-IPQDLVGDLASVVFGSQRPLLDSVAQQQGAWLPHVADFRWRVDVAISTSALARS 171

Db    145 YRPGYLVTLNVENNDSQSYPEINFSCNMEQLQDLVG-KLKDASKSLER 191
          :|: |: |:: :  :  : |:  : : ::|:  |:  ||: :  ||:
Qy    172 LQPSVLMQLKLSDGSAYRF-EVP-TAKFQELRYSVALVLKEMA-DLEK 216


RESULT    2
ID   US-08-822-260-1     STANDARD;     PRT;   195 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 1, Application US/08822260
XX
CC    Sequence 1, Application US/08822260
CC    Patent No. 5830660
CC     GENERAL INFORMATION:
CC       APPLICANT:  Hillman, Jennifer L.
CC       APPLICANT:  Goli, Surya K.
CC       TITLE OF INVENTION:  NOVEL TUMORIGENESIS PROTEIN
CC       NUMBER OF SEQUENCES:  3
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Incyte Pharmaceuticals, Inc.
CC         STREET:  3174 Porter Drive
CC         CITY:  Palo Alto
CC         STATE:  CA
CC         COUNTRY:  USA
CC         ZIP:  94304
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette
CC         COMPUTER:  IBM Compatible
CC         OPERATING SYSTEM:  DOS
CC         SOFTWARE:  FastSEQ for Windows Version 2.0
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/822,260
CC         FILING DATE:  Herewith
CC         CLASSIFICATION:  514
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:
CC         FILING DATE:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Billings, Lucy J.
CC         REGISTRATION NUMBER:  36,749
CC         REFERENCE/DOCKET NUMBER:  PF-0247 US
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  415-855-0555
CC         TELEFAX:  415-845-4166
CC         TELEX:
CC      INFORMATION FOR SEQ ID NO:  1:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  195 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       IMMEDIATE SOURCE:
CC         LIBRARY:  UTRSNOT02
CC         CLONE:  2267574
SQ   SEQUENCE   195 AA; 22151 MW; 196728 CN;

  Query Match            7.6%;  Score 122;  DB 2;  Length 195;
  Best Local Similarity  24.8%;  Pred. No. 3.27e-01;
  Matches    27;  Conservative   35;  Mismatches  38;  Indels   9;  Gaps   8;

Db     87 LED-CKFDRERI-ELFCTEYQNNKNSLEILLGSIGRSLPHITDVSWRLEYQIKTNQLHRM 144
          |:: | : :: : :|    : : :  |: :    |  |||::|  ||::  | |: | |
Qy    113 LQELC-IPQDLVGDLASVVFGSQRPLLDSVAQQQGAWLPHVADFRWRVDVAISTSALARS 171

Db    145 YRPAYLVTLSVQNTDSPSYP-EISFSCSMEQLQDLVG-KLKDASKSLER 191
          :|: |: | :  :|:::|  |:: :   ::|:  |:  ||: :  ||:
Qy    172 LQPSVLMQLKL--SDGSAYRFEVP-TAKFQELRYSVALVLKEMA-DLEK 216


RESULT    3
ID   PCT-US95-04682-2     STANDARD;     PRT;   2414 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application PC/TUS9504682
XX
CC    Sequence 2, Application PC/TUS9504682
CC     GENERAL INFORMATION:
CC       APPLICANT:
CC       TITLE OF INVENTION:  NUCLEIC ACID ENCODING TRANSCRIPTION
CC       TITLE OF INVENTION:  FACTOR P300 AND USES OF P300
CC       NUMBER OF SEQUENCES:  13
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Weingarten, Schurgin, Gagnebin & Hayes
CC         STREET:  Ten Post Office Square
CC         CITY:  Boston
CC         STATE:  MA
CC         COUNTRY:  US
CC         ZIP:  02109
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/US95/04682
CC         FILING DATE:
CC         CLASSIFICATION:
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 08/227,536
CC         FILING DATE:  14-April-1994
CC         CLASSIFICATION:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Holliday C. Heine, Ph.D.
CC         REGISTRATION NUMBER:  34,346
CC         REFERENCE/DOCKET NUMBER:  DFCI-308Xq999
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (617) 542-2290
CC         TELEFAX:  (617) 451-0313
CC      INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  2414 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   2414 AA; 264143 MW; 29411911 CN;

  Query Match            6.9%;  Score 110;  DB 3;  Length 2414;
  Best Local Similarity  34.0%;  Pred. No. 2.22e+00;
  Matches    17;  Conservative   13;  Mismatches  18;  Indels   2;  Gaps   2;

Db   2137 QAGVQRAGLPQQQPQQQLQPPMGGMSPQAQQ-MNMNHNTM-PSQFRDILR 2184
          : |||| |::   |::|| : ::||    || : :   :: |  ||| |:
Qy     65 RDGVQRLGVSANLPEEQLGALLAGMHTLLQQALRLPPTSLKPDTFRDQLQ 114
```

```
RESULT    4
ID   US-08-227-536-2     STANDARD;     PRT;   2414 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application US/08227536
XX
CC    Sequence 2, Application US/08227536
CC    Patent No. 5658784
CC     GENERAL INFORMATION:
CC       APPLICANT:  Eckner, Richard
CC       APPLICANT:  Ewen, Mark
CC       APPLICANT:  Livingston, David
CC       TITLE OF INVENTION:  NUCLEIC ACID, ENCODING TRANSCRIPTION
CC       TITLE OF INVENTION:  FACTOR P300 AND USES OF P300
CC       NUMBER OF SEQUENCES:  13
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Weingarten, Schurgin, Gagnebin & Hayes
CC         STREET:  Ten Post Office Square
CC         CITY:  Boston
CC         STATE:  MA
CC         COUNTRY:  US
CC         ZIP:  02109
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/227,536
CC         FILING DATE:  14-APR-1994
CC         CLASSIFICATION:  436
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Williams Ph.D., Kathleen A.
CC         REGISTRATION NUMBER:  34,380
CC         REFERENCE/DOCKET NUMBER:  DFCI-308XX
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (617) 542-2290
CC         TELEFAX:  (617) 451-0313
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  2414 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   2414 AA; 264143 MW; 29411911 CN;

  Query Match            6.9%;  Score 110;  DB 1;  Length 2414;
  Best Local Similarity  34.0%;  Pred. No. 2.22e+00;
  Matches    17;  Conservative   13;  Mismatches  18;  Indels   2;  Gaps   2;

Db   2137 QAGVQRAGLPQQQPQQQLQPPMGGMSPQAQQ-MNMNHNTM-PSQFRDILR 2184
           : |||| |::   |::|| : ::||    || : :   :: |  ||| |:
Qy     65 RDGVQRLGVSANLPEEQLGALLAGMHTLLQQALRLPPTSLKPDTFRDQLQ 114


RESULT    5
ID   US-08-795-303-18     STANDARD;     PRT;   395 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 18, Application US/08795303
XX
CC    Sequence 18, Application US/08795303
CC    Patent No. 5948656
CC     GENERAL INFORMATION:
CC       APPLICANT:  Anderson, Paul J.
CC       APPLICANT:  Tian, Qingsheng
CC       TITLE OF INVENTION:  TIA-1 BINDING PROTEINS AND ISOLATED
CC       TITLE OF INVENTION:  COMPLEMENTARY DNA ENCODING THE SAME
CC       NUMBER OF SEQUENCES:  21
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Sughrue, Mion, Zinn, Macpeak & Seas
CC         STREET:  2100 Pennsylvania Avenue, NW  Suite 800
CC         CITY:  Washington
CC         STATE:  DC
CC         COUNTRY:  USA
CC         ZIP:  20037
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/795,303
CC         FILING DATE:  04-FEB-1997
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  08/318,947
CC         FILING DATE:  06-OCT-1994
CC         APPLICATION NUMBER:  08/133,530
CC         FILING DATE:  07-OCT-1993
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Mack, Susan J.
CC         REGISTRATION NUMBER:  30,951
CC         REFERENCE/DOCKET NUMBER:  A6462
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (202)293-7060
CC         TELEFAX:  (202)293-2920
CC         TELEX:  6491103
CC     INFORMATION FOR SEQ ID NO:  18:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  395 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   395 AA; 42522 MW; 753192 CN;

  Query Match            6.7%;  Score 107;  DB 2;  Length 395;
  Best Local Similarity  29.7%;  Pred. No. 3.56e+00;
  Matches    27;  Conservative   25;  Mismatches  33;  Indels   6;  Gaps   5;

Db    180 ELTPENAEAVARFLGDAVDREP-ALMLEYFCRCAREESKRVPPRTFGSAPRLTEDDFGLL 238
          :| || | |:||:||| :||     :| :     : |  |   : :| :: | |:::| |
Qy     28 QLPPEVA-AMARLLGD-LDRSTFRKLLKFVVSSLQGEDCRDGVQRLGVSANLPEEQLGAL 85

Db    239 NTALAEMRRLCLDLPPVP--PNAYTPYHLRE 267
          :::  : :  | ||| :  |:::   :|:|
Qy     86 LAGMHTLLQQALRLPPTSLKPDTFRD-QLQE 115


RESULT    6
ID   US-08-318-947A-18     STANDARD;     PRT;   395 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 18, Application US/08318947A
XX
CC    Sequence 18, Application US/08318947A
CC    Patent No. 5798245
CC     GENERAL INFORMATION:
CC       APPLICANT:  Anderson, Paul J.
CC       APPLICANT:  Tian, Qingsheng
CC       TITLE OF INVENTION:  TIA-1 BINDING PROTEINS AND ISOLATED
CC       TITLE OF INVENTION:  COMPLEMENTARY DNA ENCODING THE SAME
CC       NUMBER OF SEQUENCES:  21
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Sughrue, Mion, Zinn, Macpeak & Seas
CC         STREET:  2100 Pennsylvania Avenue, NW  Suite 800
CC         CITY:  Washington
```

```
CC         STATE:  DC
CC         COUNTRY:  USA
CC         ZIP:  20037
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/318,947A
CC         FILING DATE:  06-OCT-1994
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  08/133,530
CC         FILING DATE:  07-OCT-1993
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Mack, Susan J.
CC         REGISTRATION NUMBER:  30,951
CC         REFERENCE/DOCKET NUMBER:  A6462
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (202)293-7060
CC         TELEFAX:  (202)293-2920
CC         TELEX:  6491103
CC     INFORMATION FOR SEQ ID NO:  18:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  395 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   395 AA; 42522 MW; 753192 CN;

  Query Match            6.7%;  Score 107;  DB 1;  Length 395;
  Best Local Similarity  29.7%;  Pred. No. 3.56e+00;
  Matches    27;  Conservative   25;  Mismatches  33;  Indels   6;  Gaps   5;

Db    180 ELTPENAEAVARFLGDAVDREP-ALMLEYFCRCAREESKRVPPRTFGSAPRLTEDDFGLL 238
          :| || | |:||:||| :||     :| :     : |  |   : :| :: | |:::| |
Qy     28 QLPPEVA-AMARLLGD-LDRSTFRKLLKFVVSSLQGEDCRDGVQRLGVSANLPEEQLGAL 85

Db    239 NTALAEMRRLCLDLPPVP--PNAYTPYHLRE 267
          :::  : :  | ||| :  |:::   :|:|
Qy     86 LAGMHTLLQQALRLPPTSLKPDTFRD-QLQE 115


RESULT    7
ID   US-08-576-626A-31     STANDARD;     PRT;   1114 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 31, Application US/08576626A
XX
CC    Sequence 31, Application US/08576626A
CC    Patent No. 5998194
CC     GENERAL INFORMATION:
CC       APPLICANT:  Summers, R.G.
CC       APPLICANT:  Katz, L.
CC       APPLICANT:  Donadio, S.
CC       APPLICANT:  Staver, M.J.
CC       TITLE OF INVENTION:  POLYKETIDE-ASSOCIATED SUGAR
CC       TITLE OF INVENTION:  BIOSYNTHESIS GENES
CC       NUMBER OF SEQUENCES:  60
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Abbott Laboratories
CC         STREET:  100 Abbott Park Road
CC         CITY:  Abbott Park
CC         STATE:  Illinois
CC         COUNTRY:  USA
CC         ZIP:  60064-3500
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette
CC         COMPUTER:  IBM Compatible
CC         OPERATING SYSTEM:  DOS
CC         SOFTWARE:  FastSEQ Version 2.0
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/576,626A
CC         FILING DATE:  21-DEC-1995
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:
CC         FILING DATE:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Dianne Casuto
CC         REGISTRATION NUMBER:  P-40,943
CC         REFERENCE/DOCKET NUMBER:  5857.US.01
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (847) 938-3137
CC         TELEFAX:  (847) 938-2623
CC         TELEX:
CC     INFORMATION FOR SEQ ID NO:  31:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1114 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  No. 5998194e
SQ   SEQUENCE   1114 AA; 120173 MW; 5916545 CN;

  Query Match            6.7%;  Score 107;  DB 2;  Length 1114;
  Best Local Similarity  25.2%;  Pred. No. 3.56e+00;
  Matches    33;  Conservative   36;  Mismatches  54;  Indels   8;  Gaps   7;

Db    150 MTAVLGAAGVLR--GAAWDARVS-LDAQLSPQQLAVTEAAVAALPADPALRAL-FAGAEM 205
          |:|| :|:  |:  | :  :||| | |||:|: :|     :: |      : || |  : :
Qy      1 MSAVGAATPYLHHPGDSHSGRVSFLGAQLPPE-VAAMARLLGDLDRSTFRKLLKFVVSSL 59

Db    206 TANTVVDAVLAVSAEPGLAE-RIADDPAAAQRTVAEVLRLHPA-LHLER-RTATAEVRLG 262
          ::   |:|  ::  : |:| :::   |: :  : : ||| |: |  :  |    |: :
Qy     60 QGEDCRDGVQRLGVSANLPEEQLGALLAGMHTLLQQALRLPPTSLKPDTFRDQLQELCIP 119

Db    263 EHVIGEGEEVV 273
          : ::|:   ||
Qy    120 QDLVGDLASVV 130


RESULT    8
ID   US-09-015-815-1     STANDARD;     PRT;   448 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 1, Application US/09015815
XX
CC    Sequence 1, Application US/09015815
CC    Patent No. 5965356
CC    GENERAL INFORMATION:
CC     APPLICANT: AURELIAN, LAURE
CC     APPLICANT: SMITH, CYNTHIA
CC     TITLE OF INVENTION: HERPES SIMPLEX VIRUS TYPE SPECIFIC SEROASSAY
CC     FILE REFERENCE: 1437LA
CC     CURRENT APPLICATION NUMBER: US/09/015,815
CC     CURRENT FILING DATE: 1998-01-29
CC     EARLIER APPLICATION NUMBER: US 60/036,622
CC     EARLIER FILING DATE: 1997-01-31
CC     NUMBER OF SEQ ID NOS: 1
CC     SOFTWARE: PatentIn Ver. 2.0
CC    SEQ ID NO 1
CC      LENGTH: 448
CC      TYPE: PRT
CC      ORGANISM: herpes simplex virus-2
CC      FEATURE:
CC      NAME/KEY: DOMAIN
CC      LOCATION: (1)..(446)
CC      OTHER INFORMATION: PROTEIN KINASE DOMAIN OF ICP10 SUBUNIT OF HSV-2
```

```
CC      PUBLICATION INFORMATION:
CC       AUTHORS: CHUNG ET AL.,
CC       JOURNAL: J. Virol.
CC       VOLUME: 63
CC       PAGES: 3389-3398
CC       DATE: 1989
CC      PUBLICATION INFORMATION:
CC       AUTHORS: NELSON ET AL.,
CC       JOURNAL: J. Biol. Chem.
CC       VOLUME: 271
CC       PAGES: 17021-17027
CC       DATE: 1996
SQ   SEQUENCE   448 AA; 46940 MW; 974341 CN;

  Query Match             6.2%;  Score 100;  DB 2;  Length 448;
  Best Local Similarity  29.7%;  Pred. No. 1.05e+01;
  Matches    27;  Conservative   24;  Mismatches  34;  Indels   6;  Gaps   5;

Db    324 ELTPENAEAVARFLGDAVDREP-ALMLEYFCRCAREESKRVPPRTFGSAPRLTEDDFGLL 382
          :| || | |:||:||| :||     :| :     : |  |   : :| :: | |:::| |
Qy     28 QLPPEVA-AMARLLGD-LDRSTFRKLLKFVVSSLQGEDCRDGVQRLGVSANLPEEQLGAL 85

Db    383 NYALAEMRRLCLDLPPVP--PNAYTPYHLRE 411
            ::  : :  | ||| :  |:::   :|:|
Qy     86 LAGMHTLLQQALRLPPTSLKPDTFRD-QLQE 115


RESULT    9
ID   US-09-003-289-2     STANDARD;     PRT;   1064 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application US/09003289
XX
CC    Sequence 2, Application US/09003289
CC    Patent No. 5916792
CC     GENERAL INFORMATION:
CC       APPLICANT:  Civin, Curt I.
CC       APPLICANT:  Small, Donald
CC       TITLE OF INVENTION:  NOVEL PROTEIN TYROSINE KINASE, JAK3
CC       NUMBER OF SEQUENCES:  12
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Fish & Richardson P.C.
CC         STREET:  4225 Executive Square, Suite 1400
CC         CITY:  La Jolla
CC         STATE:  CA
CC         COUNTRY:  USA
CC         ZIP:  92037
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/09/003,289
CC         FILING DATE:
CC         CLASSIFICATION:
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/357,598
CC         FILING DATE:  15-DEC-1994
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Haile, Lisa A.
CC         REGISTRATION NUMBER:  38,347
CC         REFERENCE/DOCKET NUMBER:  07265/033001
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  619/678-5070
CC         TELEFAX:  619/678-5099
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1064 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   1064 AA; 118717 MW; 5778568 CN;

  Query Match             6.1%;  Score 98;  DB 2;  Length 1064;
  Best Local Similarity  22.8%;  Pred. No. 1.43e+01;
  Matches    18;  Conservative   26;  Mismatches  29;  Indels   6;  Gaps   5;

Db     19 LSTEAGALHVLLPARGPGPPQRL-SFSFGDHLAEDLCVQAAKASAILPV-Y--H-SLFAL 73
          |::  :::| ||      ||  | : :| |:| ::||:     : : :| :  : :|:
Qy     82 LGALLAGMHTLLQQALRLPPTSLKPDTFRDQL-QELCIPQDLVGDLASVVFGSQRPLLDS 140

Db     74 ATEDLSCWFPRATSSPWRM 92
              ::: : |:|: :   ||:
Qy    141 VAQQQGAWLPHVADFRWRV 159


RESULT   10
ID   US-08-357-598-2     STANDARD;     PRT;   1064 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application US/08357598
XX
CC    Sequence 2, Application US/08357598
CC    Patent No. 5705625
CC     GENERAL INFORMATION:
CC       APPLICANT:  Civin, Curt I.
CC       APPLICANT:  Small, Donald
CC       TITLE OF INVENTION:  NOVEL PROTEIN TYROSINE KINASE, JAK3
CC       NUMBER OF SEQUENCES:  12
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Fish & Richardson P.C.
CC         STREET:  4225 Executive Square, Suite 1400
CC         CITY:  La Jolla
CC         STATE:  CA
CC         COUNTRY:  USA
CC         ZIP:  92037
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/357,598
CC         FILING DATE:  15-DEC-1994
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Haile, Lisa A.
CC         REGISTRATION NUMBER:  38,347
CC         REFERENCE/DOCKET NUMBER:  07265/033001
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  619/678-5070
CC         TELEFAX:  619/678-5099
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1064 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   1064 AA; 118717 MW; 5778568 CN;

  Query Match             6.1%;  Score 98;  DB 1;  Length 1064;
  Best Local Similarity  22.8%;  Pred. No. 1.43e+01;
  Matches    18;  Conservative   26;  Mismatches  29;  Indels   6;  Gaps   5;

Db     19 LSTEAGALHVLLPARGPGPPQRL-SFSFGDHLAEDLCVQAAKASAILPV-Y--H-SLFAL 73
          |::  :::| ||      ||  | : :| |:| ::||:     : : :| :  : :|:
Qy     82 LGALLAGMHTLLQQALRLPPTSLKPDTFRDQL-QELCIPQDLVGDLASVVFGSQRPLLDS 140

Db     74 ATEDLSCWFPRATSSPWRM 92
```

```
          ::: : |:|: :   ||:
Qy    141 VAQQQGAWLPHVADFRWRV 159


RESULT   11
ID   PCT-US95-16435-2     STANDARD;     PRT;   1064 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application PC/TUS9516435
XX
CC    Sequence 2, Application PC/TUS9516435
CC     GENERAL INFORMATION:
CC       APPLICANT:  The Johns Hopkins University School of Medicine
CC       TITLE OF INVENTION:  NOVEL PROTEIN TYROSINE KINASE, JAK3
CC       NUMBER OF SEQUENCES:  12
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Fish & Richardson P.C.
CC         STREET:  4225 Executive Square, Suite 1400
CC         CITY:  La Jolla
CC         STATE:  CA
CC         COUNTRY:  USA
CC         ZIP:  92037
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/US95/16435
CC         FILING DATE:  15-DEC-1995
CC         CLASSIFICATION:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Haile, Lisa A.
CC         REGISTRATION NUMBER:  38,347
CC         REFERENCE/DOCKET NUMBER:  07265/033WO1
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  619/678-5070
CC         TELEFAX:  619/678-5099
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1064 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   1064 AA; 118717 MW; 5778568 CN;

  Query Match             6.1%;  Score 98;  DB 3;  Length 1064;
  Best Local Similarity  22.8%;  Pred. No. 1.43e+01;
  Matches    18;  Conservative   26;  Mismatches  29;  Indels   6;  Gaps   5;

Db     19 LSTEAGALHVLLPARGPGPPQRL-SFSFGDHLAEDLCVQAAKASAILPV-Y--H-SLFAL 73
          |::  :::| ||      ||  | : :| |:| ::||:     : : :| :  : :|:
Qy     82 LGALLAGMHTLLQQALRLPPTSLKPDTFRDQL-QELCIPQDLVGDLASVVFGSQRPLLDS 140

Db     74 ATEDLSCWFPRATSSPWRM 92
          ::: : |:|: :   ||:
Qy    141 VAQQQGAWLPHVADFRWRV 159


RESULT   12
ID   US-09-003-289-5     STANDARD;     PRT;   1082 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 5, Application US/09003289
XX
CC    Sequence 5, Application US/09003289
CC    Patent No. 5916792
CC     GENERAL INFORMATION:
CC       APPLICANT:  Civin, Curt I.
CC       APPLICANT:  Small, Donald
CC       TITLE OF INVENTION:  NOVEL PROTEIN TYROSINE KINASE, JAK3
CC       NUMBER OF SEQUENCES:  12
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Fish & Richardson P.C.
CC         STREET:  4225 Executive Square, Suite 1400
CC         CITY:  La Jolla
CC         STATE:  CA
CC         COUNTRY:  USA
CC         ZIP:  92037
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/09/003,289
CC         FILING DATE:
CC         CLASSIFICATION:
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/357,598
CC         FILING DATE:  15-DEC-1994
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Haile, Lisa A.
CC         REGISTRATION NUMBER:  38,347
CC         REFERENCE/DOCKET NUMBER:  07265/033001
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  619/678-5070
CC         TELEFAX:  619/678-5099
CC     INFORMATION FOR SEQ ID NO:  5:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1082 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  No. 5916792 Relevant
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   1082 AA; 120714 MW; 5959860 CN;

  Query Match             6.1%;  Score 98;  DB 2;  Length 1082;
  Best Local Similarity  22.8%;  Pred. No. 1.43e+01;
  Matches    18;  Conservative   26;  Mismatches  29;  Indels   6;  Gaps   5;

Db     19 LSTEAGALHVLLPARGPGPPQRL-SFSFGDHLAEDLCVQAAKASAILPV-Y--H-SLFAL 73
          |::  :::| ||      ||  | : :| |:| ::||:     : : :| :  : :|:
Qy     82 LGALLAGMHTLLQQALRLPPTSLKPDTFRDQL-QELCIPQDLVGDLASVVFGSQRPLLDS 140

Db     74 ATEDLSCWFPRATSSPWRM 92
          ::: : |:|: :   ||:
Qy    141 VAQQQGAWLPHVADFRWRV 159


RESULT   13
ID   US-08-357-598-10     STANDARD;     PRT;   1082 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 10, Application US/08357598
XX
CC    Sequence 10, Application US/08357598
CC    Patent No. 5705625
CC     GENERAL INFORMATION:
CC       APPLICANT:  Civin, Curt I.
CC       APPLICANT:  Small, Donald
CC       TITLE OF INVENTION:  NOVEL PROTEIN TYROSINE KINASE, JAK3
CC       NUMBER OF SEQUENCES:  12
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Fish & Richardson P.C.
CC         STREET:  4225 Executive Square, Suite 1400
CC         CITY:  La Jolla
```

```
CC         STATE:  CA
CC         COUNTRY:  USA
CC         ZIP:  92037
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/357,598
CC         FILING DATE:  15-DEC-1994
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Haile, Lisa A.
CC         REGISTRATION NUMBER:  38,347
CC         REFERENCE/DOCKET NUMBER:  07265/033001
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  619/678-5070
CC         TELEFAX:  619/678-5099
CC     INFORMATION FOR SEQ ID NO:  10:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1082 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  not relevant
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   1082 AA; 120714 MW; 5959860 CN;

  Query Match             6.1%;  Score 98;  DB 1;  Length 1082;
  Best Local Similarity  22.8%;  Pred. No. 1.43e+01;
  Matches    18;  Conservative   26;  Mismatches  29;  Indels   6;  Gaps   5;

Db      19 LSTEAGALHVLLPARGPGPPQRL-SFSFGDHLAEDLCVQAAKASAILPV-Y--H-SLFAL 73
           |::  :::| ||      ||  | : :| |:| ::||:      : : :| :  : :|:
Qy      82 LGALLAGMHTLLQQALRLPPTSLKPDTFRDQL-QELCIPQDLVGDLASVVFGSQRPLLDS 140

Db      74 ATEDLSCWFPRATSSPWRM 92
           ::: : |:|: :   ||:
Qy     141 VAQQQGAWLPHVADFRWRV 159


RESULT   14
ID   PCT-US95-16435-5     STANDARD;     PRT;   1082 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 5, Application PC/TUS9516435
XX
CC    Sequence 5, Application PC/TUS9516435
CC     GENERAL INFORMATION:
CC       APPLICANT:  The Johns Hopkins University School of Medicine
CC       TITLE OF INVENTION:  NOVEL PROTEIN TYROSINE KINASE, JAK3
CC       NUMBER OF SEQUENCES:  12
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Fish & Richardson P.C.
CC         STREET:  4225 Executive Square, Suite 1400
CC         CITY:  La Jolla
CC         STATE:  CA
CC         COUNTRY:  USA
CC         ZIP:  92037
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/US95/16435
CC         FILING DATE:  15-DEC-1995
CC         CLASSIFICATION:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Haile, Lisa A.
CC         REGISTRATION NUMBER:  38,347
CC         REFERENCE/DOCKET NUMBER:  07265/033WO1
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  619/678-5070
CC         TELEFAX:  619/678-5099
CC     INFORMATION FOR SEQ ID NO:  5:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1082 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  not relevant
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   1082 AA; 120714 MW; 5959860 CN;

  Query Match             6.1%;  Score 98;  DB 3;  Length 1082;
  Best Local Similarity  22.8%;  Pred. No. 1.43e+01;
  Matches    18;  Conservative   26;  Mismatches  29;  Indels   6;  Gaps   5;

Db      19 LSTEAGALHVLLPARGPGPPQRL-SFSFGDHLAEDLCVQAAKASAILPV-Y--H-SLFAL 73
           |::  :::| ||      ||  | : :| |:| ::||:      : : :| :  : :|:
Qy      82 LGALLAGMHTLLQQALRLPPTSLKPDTFRDQL-QELCIPQDLVGDLASVVFGSQRPLLDS 140

Db      74 ATEDLSCWFPRATSSPWRM 92
           ::: : |:|: :   ||:
Qy     141 VAQQQGAWLPHVADFRWRV 159


RESULT   15
ID   PCT-US95-16435-10     STANDARD;     PRT;   1082 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 10, Application PC/TUS9516435
XX
CC    Sequence 10, Application PC/TUS9516435
CC     GENERAL INFORMATION:
CC       APPLICANT:  The Johns Hopkins University School of Medicine
CC       TITLE OF INVENTION:  NOVEL PROTEIN TYROSINE KINASE, JAK3
CC       NUMBER OF SEQUENCES:  12
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Fish & Richardson P.C.
CC         STREET:  4225 Executive Square, Suite 1400
CC         CITY:  La Jolla
CC         STATE:  CA
CC         COUNTRY:  USA
CC         ZIP:  92037
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/US95/16435
CC         FILING DATE:  15-DEC-1995
CC         CLASSIFICATION:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Haile, Lisa A.
CC         REGISTRATION NUMBER:  38,347
CC         REFERENCE/DOCKET NUMBER:  07265/033WO1
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  619/678-5070
CC         TELEFAX:  619/678-5099
CC     INFORMATION FOR SEQ ID NO:  10:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1082 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  not relevant
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   1082 AA; 120714 MW; 5959860 CN;
```

```
  Query Match           6.1%;  Score 98;  DB 3;  Length 1082;
  Best Local Similarity  22.8%;  Pred. No. 1.43e+01;
  Matches   18;  Conservative   26;  Mismatches  29;  Indels   6;  Gaps   5;

Db      19 LSTEAGALHVLLPARGPGPPQRL-SFSFGDHLAEDLCVQAAKASAILPV-Y--H-SLFAL 73
           |::  :::| ||      ||  | : :| |:| ::||:     : : :| :  : :|:
Qy      82 LGALLAGMHTLLQQALRLPPTSLKPDTFRDQL-QELCIPQDLVGDLASVVFGSQRPLLDS 140

Db      74 ATEDLSCWFPRATSSPWRM 92
           ::: : |:|: :   ||:
Qy     141 VAQQQGAWLPHVADFRWRV 159


Search completed: Fri May 12 10:36:40 2000
Job time : 22 secs.
```

****

MPSRCH (TM)

****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 12 10:35:27 2000; MasPar time 29.93 Seconds
177.250 Million cell updates/sec
Tabular output not generated.

Title: >US-09-223-796-4
Description: (1-224) from US09223796.pep
Perfect Score: 1604
Sequence: 1 MSAVGAATPYLHHPGDSHSG.....ALVLKEMADLEKRCERRLQD 224

Scoring table: PAM 150
Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq36
1:geneseqp

Statistics: Mean 33.179; Variance 157.131; scale 0.211

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	1319	82.2	224	1	W37723	Rat Hypertension relat	1.42e-104
2	122	7.6	195	1	W71684	Amino acid sequence of	6.45e-01
3	122	7.6	196	1	Y07867	Human secreted protein	6.45e-01
4	116	7.2	169	1	W88390	Human Zneu1 partial po	1.72e+00
5	116	7.2	181	1	W88391	Human Zneu1 partial po	1.72e+00
6	116	7.2	254	1	W88382	Human neuro-growth fac	1.72e+00
7	116	7.2	273	1	W88381	Human neuro-growth fac	1.72e+00
8	110	6.9	800	1	R84883	Transcription factor p	4.50e+00
9	110	6.9	2414	1	R84882	Transcription factor p	4.50e+00
10	110	6.9	2414	1	W40057	Cellular transcription	4.50e+00
11	107	6.7	360	1	W19734	Sugar biosynthesis enz	7.24e+00
12	107	6.7	361	1	W99387	S.erythraea dTDP-4-ket	7.24e+00
13	106	6.6	339	1	R77317	Protein activated lipa	8.48e+00
14	106	6.6	727	1	W99798	Human VRRP-1 (VR2) cap	8.48e+00
15	104	6.5	344	1	R47213	Lipase modulator.	1.16e+01
16	100	6.2	694	1	R04107	DNA-binding protein GC	2.16e+01
17	100	6.2	784	1	W34179	Human GC binding prote	2.16e+01
18	100	6.2	1144	1	W72206	HSV-2 strain SB5 Conti	2.16e+01
19	100	6.2	1180	1	W72020	HSV-2 strain SB5 Conti	2.16e+01
20	100	6.2	1180	1	W72098	HSV-2 strain SB5 Conti	2.16e+01
21	98	6.1	1064	1	R96037	Protein tyrosine kinas	2.94e+01
22	95	5.9	603	1	R85889	WD-40 domain-contg. ra	4.65e+01
23	94	5.9	1137	1	P71182	Sequence of herpes sim	5.41e+01
24	93	5.8	1291	1	R68975	Entamoeba histolytica	6.29e+01
25	93	5.8	1292	1	R70681	Entamoeba histolytica	6.29e+01
26	92	5.7	335	1	R99975	Pseudomonas sp. SD705	7.31e+01
27	92	5.7	350	1	W60619	Human liver activin be	7.31e+01
28	92	5.7	350	1	R92754	Human growth different	7.31e+01
29	91	5.7	476	1	W02616	Wheat adenylosuccinate	8.49e+01
30	91	5.7	490	1	W99454	Arabidopsis adenylosuc	8.49e+01
31	91	5.7	490	1	R97733	Arabidopsis adenylosuc	8.49e+01
32	91	5.7	490	1	W02614	Arabidopsis adenylosuc	8.49e+01
33	91	5.7	822	1	W41943	Human OC-116 kDa prote	8.49e+01
34	91	5.7	1319	1	R45002	Cellulose synthase ope	8.49e+01
35	91	5.7	1326	1	W69758	Acetobacter xylinum bc	8.49e+01
36	91	5.7	1336	1	W64232	Seq ID #8 from DE19653	8.49e+01
37	90	5.6	332	1	W52284	Mutant prenyl diphosph	9.86e+01
38	90	5.6	561	1	W63701	Human hSK1 protein.	9.86e+01
39	89	5.5	182	1	W20289	H. pylori transporter	1.14e+02
40	88	5.5	279	1	W74908	Human secreted protein	1.33e+02
41	89	5.5	494	1	R31888	Defective tyrosine kin	1.14e+02
42	89	5.5	494	1	R76615	Saccharomyces cerevisi	1.14e+02
43	89	5.5	494	1	R56519	Protein kinase (HRR25)	1.14e+02
44	89	5.5	519	1	W20640	H. pylori transporter	1.14e+02
45	88	5.5	686	1	R25591	RING11 antigenic pepti	1.33e+02

ALIGNMENTS

```
RESULT    1
ID   W37723 standard; Protein; 224 AA.
AC   W37723;
DT   09-JUN-1998  (first entry)
DE   Rat Hypertension related calcium regulator.
KW   Hypertension related calcium regulated gene; HCaRG; rat parathyroid;
KW   extracellular calcium concentration; antibody; hypertension;
KW   hyperthyroidism; osteoporosis; heart failure; diabetes; stroke;
KW   cancer; inflammatory disease; asthma.
OS   Rattus rattus.
FH   Key             Location/Qualifiers
FT   Misc_difference 15. .21
FT                   /note= "EF-hand like motif"
PN   WO9749807-A2.
PD   31-DEC-1997.
PF   23-JUN-1997; CA0439.
PR   21-JUN-1996; US-667495.
PA   (GOSS/) GOSSARD F.
PA   (HAME/) HAMET P.
PA   (LEWA/) LEWANCZUK R.
PA   (TREM/) TREMBLAY J.
PI   Gossard F, Hamet P, Lewanczuk R, Tremblay J;
DR   WPI; 98-077171/07.
DR   N-PSDB; V18890.
PT   Hypertension related calcium regulated gene - useful to develop
PT   products to treat or detect, e.g. hypertension, stroke,
PT   osteoporosis, heart failure, cancer, diabetes or asthma
PS   Claim 8; Pages 26-27; 46pp; English.
CC   This is the amino acid sequence of the hypertension related calcium
CC   regulated gene (HCaRG), which was isolated from the rat parathyroid.
CC   Its expression is regulated by extracellular calcium concentration.
CC   An antibody against the protein, can be used to detect or modulate
CC   (e.g. enhance or inhibit) abnormal calcium levels.  They can
CC   specifically be used to detect or treat, e.g. hypertension,
CC   hyperthyroidism, osteoporosis, heart failure, diabetes, stroke, cancer,
CC   inflammatory disease, and asthma.
SQ   Sequence   224 AA;

  Query Match          82.2%;  Score 1319;  DB 1;  Length 224;
  Best Local Similarity  79.5%;  Pred. No. 1.42e-104;
  Matches   178;  Conservative   33;  Mismatches  13;  Indels   0;  Gaps   0;

Db       1 MSALGAAAPYLHHPADSHSGRVSFLGSQPSPEVTAVAQLLKDLDRSTFRKLLKLVVGALH 60
           |||:|||:||||||:||||||||||:| :|||:|:|:|| ||||||||||||:||::|:
Qy       1 MSAVGAATPYLHHPGDSHSGRVSFLGAQLPPEVAAMARLLGDLDRSTFRKLLKFVVSSLQ 60

Db      61 GKDCREAVEQLGASANLSEERLAVLLAGTHTLLQQALRLPPASLKPDAFQEELQELGIPQ 120
```

```
          | |||::|::|| ||||:||:|: |||| |||||||||||:|||||:|:::|||| |||
Qy      61 GEDCRDGVQRLGVSANLPEEQLGALLAGMHTLLQQALRLPPTSLKPDTFRDQLQELCIPQ 120

Db     121 DLIGDLASLAFGSQRPLLDSVAQQQGSSLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQL 180
           ||:|||||: |||||||||||||||: ||||: |||||||||||| :||||||||||
Qy     121 DLVGDLASVVFGSQRPLLDSVAQQQGAWLPHVADFRWRVDVAISTSALARSLQPSVLMQL 180

Db     181 KLTDGSAHRFEVPIAKFQELRYSVALVLKEMAELEKKCERKLQD 224
           ||:|||| ||||| ||||||||||||||||:|||:|||:|||
Qy     181 KLSDGSAYRFEVPTAKFQELRYSVALVLKEMADLEKRCERRLQD 224


RESULT    2
ID   W71684 standard; Protein; 195 AA.
AC   W71684;
DT   04-DEC-1998  (first entry)
DE   Amino acid sequence of the human tumourigenesis associated protein.
KW   Human; tumourigenesis associated protein; HTAP; transplantation;
KW   tumour; Antagonist; cancer; inflammation; immunological disease;
KW   antibody; probe; primer; PCR; amplification; hybridisation;
KW   inhibition.
OS   Homo sapiens.
PN   WO9841635-A1.
PD   24-SEP-1998.
PF   20-MAR-1998; U06066.
PR   20-MAR-1997; US-822260.
PA   (INCY-) INCYTE PHARM INC.
PI   Goli SK, Hillman JL;
DR   WPI; 98-521224/44.
DR   N-PSDB; V58281.
PT   New tumorigenesis-associated protein and related nucleic acid,
PT   vectors, transformed cells - antibodies, agonists and antagonists,
PT   for diagnosis, treatment and prevention of abnormal cellular
PT   differentiation, particularly cancers and inflammation
PS   Claim 1; Fig 1A-1B; 54pp; English.
CC   This is the amino acid sequence of the human tumourigenesis
CC   associated protein (HTAP), used in the method of the invention.  HTAP,
CC   is involved with cell proliferation and inflammation.  It can be used
CC   to stimulate cell proliferation (e.g. of cells intended for
CC   transplantation in treatment of tumours or infections, or to treat
CC   genetic defects).  Antagonists of HTAP are used to treat or prevent
CC   a wide range of cancers (adenocarcinoma, melanoma, sarcoma, lymphoma,
CC   leukaemia etc.), also inflammation where associated with infection or
CC   immunological disease (e.g. asthma, cystic fibrosis, rheumatoid
CC   arthritis).  HTAP is also used to raise antibodies and to screen
CC   libraries for specific-binding agents. The antibodies are used as for
CC   diagnosis or monitoring of HTAP-related diseases (in usual
CC   immunoassays), in competitive drug screens and to isolate HTAP from
CC   its natural sources.  HTAP derived probes or primers, are used in
CC   standard amplification or hybridisation tests to diagnose HTAP-related
CC   diseases; to identify related sequences; for genomic mapping and for
CC   screening for specific inhibitors.
SQ   Sequence   195 AA;

  Query Match           7.6%;  Score 122;  DB 1;  Length 195;
  Best Local Similarity  24.8%;  Pred. No. 6.45e-01;
  Matches   27;  Conservative   35;  Mismatches  38;  Indels   9;  Gaps   8;

Db      87 LED-CKFDRERI-ELFCTEYQNNKNSLEILLGSIGRSLPHITDVSWRLEYQIKTNQLHRM 144
           |:: | : :: : :|    : : :  |: :    |  |||::|  ||::  | |: | |
Qy     113 LQELC-IPQDLVGDLASVVFGSQRPLLDSVAQQQGAWLPHVADFRWRVDVAISTSALARS 171

Db     145 YRPAYLVTLSVQNTDSPSYP-EISFSCSMEQLQDLVG-KLKDASKSLER 191
             :|: |: | :  :|:::|  |:: :   ::|:  |:  ||: :  ||:
Qy     172 LQPSVLMQLKL--SDGSAYRFEVP-TAKFQELRYSVALVLKEMA-DLEK 216


RESULT    3
ID   Y07867 standard; Protein; 196 AA.
AC   Y07867;
DT   06-JUL-1999  (first entry)
DE   Human secreted protein fragment encoded from gene 16.
KW   Human; secreted protein; treatment; prevention; protein therapy; AIDS;
KW   gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
KW   developmental abnormality; fetal deficiency; blood disorder; leukemia;
KW   immune system disease; autoimmune disease; hepatic disease; lymphoma;
KW   renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;
KW   cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;
KW   pulmonary disorder; transplant rejection; osteoclast; osteoporosis;
KW   arthritis; malignancy; digestive; endocrine; infection.
OS   Homo sapiens.
PN   WO9918208-A1.
PD   15-APR-1999.
PF   01-OCT-1998; U20775.
PR   02-OCT-1997; US-060884.
PR   02-OCT-1997; US-060833.
PR   02-OCT-1997; US-060836.
PR   02-OCT-1997; US-060837.
PR   02-OCT-1997; US-060838.
PR   02-OCT-1997; US-060839.
PR   02-OCT-1997; US-060843.
PR   02-OCT-1997; US-060862.
PR   02-OCT-1997; US-060866.
PR   02-OCT-1997; US-060874.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Carter KC, Duan DR, Endress GA, Feng P, Ferrie AM,
PI   Florence KA, Greene JM, Janat F, Lafleur DW, Ni J,
PI   Rosen CA, Ruben SM, Shi Y, Young P, Yu G;
DR   WPI; 99-264022/22.
DR   N-PSDB; X37466.
PT   New isolated human genes and the secreted polypeptides they encode
PS   Claim 1b; Page 285-286; 368pp; English.
CC   This invention describes novel isolated human genes and the secreted
CC   proteins they encode. The products of the invention are useful for
CC   preventing, treating or ameliorating medical conditions, e.g. by protein
CC   or gene therapy. Also pathological conditions can be diagnosed by
CC   determining the amount of the new polypeptides in a sample or by
CC   determining the presence of mutations in the new polynucleotides.
CC   Specific uses are described for each of the 101 polynucleotides, based on
CC   which tissues they are most highly expressed in, and include developing
CC   products for the diagnosis or treatment of cancer, tumours,
CC   neurodegenerative disorders, developmental abnormalities and fetal
CC   deficiencies, blood disorders, leukemias, diseases of the immune system,
CC   autoimmune diseases, hepatic and renal disease, lymphomas, inflammation,
CC   allergies, Alzheimer's and cognitive disorders, schizophrenia, prostate
CC   disease, skeletal or cardiac muscle disorders, pulmonary disorders,
CC   transplant rejection, disorders involving osteoclasts such as
CC   osteoporosis, arthritis or malignancies, digestive/endocrine disorders,
CC   infections and AIDS. The human secreted proteins of the invention are
CC   represented in Y07852-Y07993 and the encoding nucleic acids are
CC   represented in X37451-X37552.
SQ   Sequence   196 AA;

  Query Match           7.6%;  Score 122;  DB 1;  Length 196;
  Best Local Similarity  24.8%;  Pred. No. 6.45e-01;
  Matches   27;  Conservative   35;  Mismatches  38;  Indels   9;  Gaps   8;

Db      87 LED-CKFDRERI-ELFCTEYQNNKNSLEILLGSIGRSLPHITDVSWRLEYQIKTNQLHRM 144
           |:: | : :: : :|    : : :  |: :    |  |||::|  ||::  | |: | |
Qy     113 LQELC-IPQDLVGDLASVVFGSQRPLLDSVAQQQGAWLPHVADFRWRVDVAISTSALARS 171

Db     145 YRPAYLVTLSVQNTDSPSYP-EISFSCSMEQLQDLVG-KLKDASKSLER 191
             :|: |: | :  :|:::|  |:: :   ::|:  |:  ||: :  ||:
Qy     172 LQPSVLMQLKL--SDGSAYRFEVP-TAKFQELRYSVALVLKEMA-DLEK 216


RESULT    4
ID   W88390 standard; Protein; 169 AA.
AC   W88390;
DT   26-APR-1999  (first entry)
DE   Human Zneu1 partial polypeptide.
KW   Zneu-1; neuro-growth factor-like protein; human; breast cancer;
KW   glioblastoma; pituitary adenoma; Alzheimer's disease; therapy;
KW   nerve regeneration; haematopoiesis; fertility; contraception;
KW   antibody.
OS   Homo sapiens.
```

```
PN   WO9857983-A2.
PD   23-DEC-1998.
PF   18-JUN-1998; U12763.
PR   18-JUN-1997; US-878322.
PR   18-JUN-1997; US-050143.
PA   (ZYMO ) ZYMOGENETICS INC.
PI   Blumberg H, Jelinek LJ, Lehner JM, Sheppard PO,
PI   Whitmore TE;
DR   WPI; 99-095324/08.
PT   New mammalian Zneu1 polypeptides - used to, e.g. treat Alzheimer's
PT   disease, cancer and to repopulate blood cells
PS   Claim 6; Page 57-58; 70pp; English.
CC   This claimed polypeptide comprises amino acid residues 105-273 of
CC   the human Zneu1 mature protein (see W88382), i.e. the 2 epidermal
CC   growth factor-like domains (see also W88384-85) and the second
CC   HSMHC3W5A homology domain of Zneu1.  Zneu1 is a new neuro-growth
CC   factor-like protein that can be used as a growth, maintenance, or
CC   differentiation factor in the spinal cord, heart, spleen, testis,
CC   thyroid and lymph nodes.  Zneu1 may also play a role in breast
CC   cancer, glioblastomas, and pituitary adenomas.  Zneu1 can be used
CC   to treat Alzheimer's disease, cancer, to repopulate blood cells
CC   after chemotherapy, to stimulate myofibroblast proliferation,
CC   stimulate or inhibit growth factors made in the placenta, in
CC   fertility and contraception, or to regenerate nerves.  Claimed
CC   Zneu1 polypeptides (see also W88382-97), including specific domains
CC   of Zneu1 and epitope-bearing portions of Zneu1, can be used to
CC   raise specific antibodies for use e.g. in diagnostic assays.
SQ   Sequence   169 AA;

  Query Match             7.2%;  Score 116;  DB 1;  Length 169;
  Best Local Similarity  44.2%;  Pred. No. 1.72e+00;
  Matches    23;  Conservative    9;  Mismatches  16;  Indels   4;  Gaps   4;

Db      93 KEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLLVHSFQ-QL 143
           :: ||||    :| || |  :|| :|:|  ||| : || | ||   :|: ||
Qy      65 RDGVQRLGVSANLPEEQLGALLAGMHTLLQQAL-R-LP-PTSLKPDTFRDQL 113


RESULT    5
ID   W88391 standard; Protein; 181 AA.
AC   W88391;
DT   26-APR-1999  (first entry)
DE   Human Zneu1 partial polypeptide.
KW   Zneu-1; neuro-growth factor-like protein; human; breast cancer;
KW   glioblastoma; pituitary adenoma; Alzheimer's disease; therapy;
KW   nerve regeneration; haematopoiesis; fertility; contraception;
KW   antibody.
OS   Homo sapiens.
PN   WO9857983-A2.
PD   23-DEC-1998.
PF   18-JUN-1998; U12763.
PR   18-JUN-1997; US-878322.
PR   18-JUN-1997; US-050143.
PA   (ZYMO ) ZYMOGENETICS INC.
PI   Blumberg H, Jelinek LJ, Lehner JM, Sheppard PO,
PI   Whitmore TE;
DR   WPI; 99-095324/08.
PT   New mammalian Zneu1 polypeptides - used to, e.g. treat Alzheimer's
PT   disease, cancer and to repopulate blood cells
PS   Claim 6; Page 58; 70pp; English.
CC   This claimed polypeptide comprises amino acid residues 1-85 and
CC   159-254 of the human Zneu1 mature protein (see W88382), i.e. the
CC   first HSMHC3W5a homology domain fused to the second HSMHC3W5a
CC   homology domain of Zneu1.  Zneu1 is a new neuro-growth factor-like
CC   protein that can be used as a growth, maintenance, or
CC   differentiation factor in the spinal cord, heart, spleen, testis,
CC   thyroid and lymph nodes.  Zneu1 may also play a role in breast
CC   cancer, glioblastomas, and pituitary adenomas.  Zneu1 can be used
CC   to treat Alzheimer's disease, cancer, to repopulate blood cells
CC   after chemotherapy, to stimulate myofibroblast proliferation,
CC   stimulate or inhibit growth factors made in the placenta, in
CC   fertility and contraception, or to regenerate nerves.  Claimed
CC   Zneu1 polypeptides (see also W88382-97), including specific domains
CC   of Zneu1 and epitope-bearing portions of Zneu1, can be used to
CC   raise specific antibodies for use e.g. in diagnostic assays.
SQ   Sequence   181 AA;

  Query Match             7.2%;  Score 116;  DB 1;  Length 181;
  Best Local Similarity  44.2%;  Pred. No. 1.72e+00;
  Matches    23;  Conservative    9;  Mismatches  16;  Indels   4;  Gaps   4;

Db     105 KEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLLVHSFQ-QL 155
           :: ||||    :| || |  :|| :|:|  ||| : || | ||   :|: ||
Qy      65 RDGVQRLGVSANLPEEQLGALLAGMHTLLQQAL-R-LP-PTSLKPDTFRDQL 113


RESULT    6
ID   W88382 standard; Protein; 254 AA.
AC   W88382;
DT   26-APR-1999  (first entry)
DE   Human neuro-growth factor-like protein Zneu1 mature polypeptide.
KW   Zneu-1; neuro-growth factor-like protein; human; breast cancer;
KW   glioblastoma; pituitary adenoma; Alzheimer's disease; therapy;
KW   nerve regeneration; haematopoiesis; fertility; contraception;
KW   antibody.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Domain          1. .85
FT                   /note= "hydrophilic domain (HSM1), homologous to
FT                         an HSMHC3W5A domain "
FT   Domain          86. .116
FT                   /note= "epidermal growth factor-like domain 1"
FT   Domain          117. .158
FT                   /note= "epidermal growth factor-like domain 2"
FT   Domain          159. .254
FT                   /note= "domain HSM2 homologous to an HSMHC3W5A
FT                         domain"
PN   WO9857983-A2.
PD   23-DEC-1998.
PF   18-JUN-1998; U12763.
PR   18-JUN-1997; US-878322.
PR   18-JUN-1997; US-050143.
PA   (ZYMO ) ZYMOGENETICS INC.
PI   Blumberg H, Jelinek LJ, Lehner JM, Sheppard PO,
PI   Whitmore TE;
DR   WPI; 99-095324/08.
PT   New mammalian Zneu1 polypeptides - used to, e.g. treat Alzheimer's
PT   disease, cancer and to repopulate blood cells
PS   Claim 6; Page 48-49; 70pp; English.
CC   This polypeptide comprises human Zneu1 mature polypeptide.  Zneu1
CC   is a new neuro-growth factor-like protein (see also W88381).  Its
CC   closest human homologue is HSMHC3W5A, a gene in the HLA class III
CC   region, which is contained in a cosmid which contains Notch 4.
CC   Zneu1 is also homologous to Notch 4 in its EGF-like domains and may
CC   be involved in EGF receptor pathways. Zneu1 is widely expressed in
CC   adult tissues, with high expression in heart, placenta, spleen,
CC   testis, thyroid, spinal cord and lymph node.  Zneu1 can be used as
CC   a growth, maintenance, or differentiation factor in the spinal
CC   cord, heart, spleen, testis, thyroid and lymph nodes.  It may
CC   play a role in breast cancer, glioblastomas, and pituitary
CC   adenomas.  Zneu1 may also be used to treat Alzheimer's disease,
CC   cancer, to repopulate blood cells after chemotherapy, to stimulate
CC   myofibroblast proliferation, stimulate or inhibit growth factors
CC   made in the placenta, in fertility and contraception, or to
CC   regenerate nerves.  Claimed Zneu1 polypeptides (see also W88382-97),
CC   including specific domains of Zneu1 and epitope-bearing portions of
CC   Zneu1, can be used to raise specific antibodies for use e.g. in
CC   diagnostic assays.
SQ   Sequence   254 AA;

  Query Match             7.2%;  Score 116;  DB 1;  Length 254;
  Best Local Similarity  44.2%;  Pred. No. 1.72e+00;
  Matches    23;  Conservative    9;  Mismatches  16;  Indels   4;  Gaps   4;

Db     178 KEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLLVHSFQ-QL 228
           :: ||||    :| || |  :|| :|:|  ||| : || | ||   :|: ||
```

```
Qy      65 RDGVQRLGVSANLPEEQLGALLAGMHTLLQQAL-R-LP-PTSLKPDTFRDQL 113


RESULT    7
ID   W88381 standard; Protein; 273 AA.
AC   W88381;
DT   26-APR-1999  (first entry)
DE   Human neuro-growth factor-like protein Zneu1.
KW   Zneu-1; neuro-growth factor-like protein; human; breast cancer;
KW   glioblastoma; pituitary adenoma; Alzheimer's disease; therapy;
KW   nerve regeneration; haematopoiesis; fertility; contraception;
KW   antibody.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Peptide         1. .19
FT                   /note= "putative signal peptide"
FT   Protein         20. .273
FT                   /note= "mature protein"
FT   Domain          20. .104
FT                   /note= "hydrophilic domain (HSM1), homologous to
FT                        an HSMHC3W5A domain "
FT   Domain          105. .135
FT                   /note= "epidermal growth factor-like domain 1"
FT   Domain          136. .177
FT                   /note= "epidermal growth factor-like domain 2"
FT   Domain          178. .273
FT                   /note= "domain HSM2 homologous to an HSMHC3W5A
FT                        domain"
PN   WO9857983-A2.
PD   23-DEC-1998.
PF   18-JUN-1998; U12763.
PR   18-JUN-1997; US-878322.
PR   18-JUN-1997; US-050143.
PA   (ZYMO ) ZYMOGENETICS INC.
PI   Blumberg H, Jelinek LJ, Lehner JM, Sheppard PO,
PI   Whitmore TE;
DR   WPI; 99-095324/08.
DR   N-PSDB; V84341.
PT   New mammalian Zneu1 polypeptides - used to, e.g. treat Alzheimer's
PT   disease, cancer and to repopulate blood cells
PS   Claim 6; Page 47-48; 70pp; English.
CC   This polypeptide comprises human Zneu1, a new neuro-growth factor-
CC   like protein.  Its amino acid sequence was deduced from the
CC   nucleotide sequence (see V84341) of a cDNA clone isolated from a
CC   brain cDNA library.  Zneu1's closest human homologue is HSMHC3W5A,
CC   a gene in the HLA class III region, which is contained in a cosmid
CC   which contains Notch 4.  Zneu1 is also homologous to Notch 4 in its
CC   EGF-like domains and may be involved in EGF receptor pathways.
CC   Zneu1 is widely expressed in adult tissues, with high expression in
CC   heart, placenta, spleen, testis, thyroid, spinal cord and lymph
CC   node.  Zneu1 polypeptide can be used as a growth, maintenance, or
CC   differentiation factor in the spinal cord, heart, spleen, testis,
CC   thyroid and lymph nodes.  It may also play a role in breast cancer,
CC   glioblastomas, and pituitary adenomas.  Zneu1 may be used to treat
CC   Alzheimer's disease, cancer, to repopulate blood cells after
CC   chemotherapy, to stimulate myofibroblast proliferation, stimulate
CC   or inhibit growth factors made in the placenta, in fertility and
CC   contraception, or to regenerate nerves.  Claimed Zneu1
CC   polypeptides (see also W88382-97), including specific domains of
CC   Zneu1 and epitope-bearing portions of Zneu1, can be used to raise
CC   specific antibodies for use e.g. in diagnostic assays.
SQ   Sequence   273 AA;

  Query Match             7.2%;  Score 116;  DB 1;  Length 273;
  Best Local Similarity  44.2%;  Pred. No. 1.72e+00;
  Matches    23;  Conservative    9;  Mismatches  16;  Indels    4;  Gaps    4;

Db     197 KEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLLVHSFQ-QL 247
           :: ||||    :| || |  :|| :|:|  ||| : || | ||   :|: ||
Qy      65 RDGVQRLGVSANLPEEQLGALLAGMHTLLQQAL-R-LP-PTSLKPDTFRDQL 113


RESULT    8
ID   R84883 standard; Protein; 800 AA.
AC   R84883;
DT   01-FEB-1996  (first entry)
DE   Transcription factor p300 C-terminal (1572-2371) region.
KW   Transcription factor; p300; adenovirus; early region 1A; E1A;
KW   cancer; diagnosis; monoclonal antibody.
OS   Homo sapiens.
PN   WO9528499-A1.
PD   26-OCT-1995.
PF   13-APR-1995; U04682.
PR   14-APR-1994; US-227536.
PA   (DAND ) DANA FARBER CANCER INST INC.
PI   Eckner R,  Ewen M,  Livingston D;
DR   WPI; 95-373813/48.
PT   Nucleic acid encoding human p300 that associates with adenovirus E1A
PT   - and related vectors, host cells and screening assays, also
PT   diagnosis of cancerous and pre:cancerous tissue by detection of
PT   mutant p300
PS   Claim 17; Page 72-78; 126pp; English.
CC   A fusion protein comprising glutathione-S-transferase and amino
CC   acids 1572-2371 (R84883) of human transcription factor p300
CC   (full sequence given in R84882) was used to produce an anti-p300
CC   monoclonal antibody (MAb) useful in identifying p300 binding
CC   proteins and for characterizing p300 in immunoassays.
SQ   Sequence   800 AA;

  Query Match             6.9%;  Score 110;  DB 1;  Length 800;
  Best Local Similarity  34.0%;  Pred. No. 4.50e+00;
  Matches    17;  Conservative   13;  Mismatches  18;  Indels    2;  Gaps    2;

Db     566 QAGVQRAGLPQQQPQQQLQPPMGGMSPQAQQ-MNMNHNTM-PSQFRDILR 613
           : |||| |::   |::|| : ::||    || : :   :: |  ||| |:
Qy      65 RDGVQRLGVSANLPEEQLGALLAGMHTLLQQALRLPPTSLKPDTFRDQLQ 114


RESULT    9
ID   R84882 standard; Protein; 2414 AA.
AC   R84882;
DT   01-FEB-1996  (first entry)
DE   Transcription factor p300.
KW   Transcription factor; p300; adenovirus; early region 1A; E1A;
KW   cancer; diagnosis.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   peptide         11. .17
FT                   /label= Nuclear_location_signal
FT   region          342. .421
FT                   /label= C/H-rich_region_1
FT                   /note= "cysteine/histidine-rich region containing
FT                   2 putative zinc finger motifs"
FT   domain          1070. .1134
FT                   /label= Bromodomain
FT   region          1162. .1461
FT                   /label= C/H-rich_region_2
FT   region          1622. .1821
FT                   /label= C/H-rich_region_3
PN   WO9528499-A1.
PD   26-OCT-1995.
PF   13-APR-1995; U04682.
PR   14-APR-1994; US-227536.
PA   (DAND ) DANA FARBER CANCER INST INC.
PI   Eckner R,  Ewen M,  Livingston D;
DR   WPI; 95-373813/48.
DR   N-PSDB; T02792.
PT   Nucleic acid encoding human p300 that associates with adenovirus E1A
PT   - and related vectors, host cells and screening assays, also
PT   diagnosis of cancerous and pre:cancerous tissue by detection of
PT   mutant p300
PS   Disclosure; Page 61-78; 126pp; English.
CC   Transcription factor p300 was isolated from human 293 cells and used
CC   to generate polyclonal antisera in mice.  These were used to screen
CC   293 cDNA libraries to isolate clones contg. overlapping inserts,
CC   which were assembled to obtain a full-length cDNA sequence (T02792)
```

```
CC   encoding a protein (R84882) of predicted mol.wt. 264.236 kDa.
CC   p300 may be produced in host cells (pref. mammalian) and used to
CC   raise monoclonal antibodies or to screen cpds. for the ability to
CC   modulate p300-dependent transcription.
SQ   Sequence   2414 AA;

  Query Match             6.9%;  Score 110;  DB 1;  Length 2414;
  Best Local Similarity  34.0%;  Pred. No. 4.50e+00;
  Matches    17;  Conservative   13;  Mismatches  18;  Indels   2;  Gaps   2;

Db   2137 QAGVQRAGLPQQQPQQQLQPPMGGMSPQAQQ-MNMNHNTM-PSQFRDILR 2184
           : |||| |::   |::|| : ::||    || : :   :: |  ||| |:
Qy     65 RDGVQRLGVSANLPEEQLGALLAGMHTLLQQALRLPPTSLKPDTFRDQLQ 114


RESULT   10
ID   W40057 standard; Protein; 2414 AA.
AC   W40057;
DT   20-JUL-1998  (first entry)
DE   Cellular transcriptional factor p300.
KW   Cellular transcriptional factor; p300; human; P300; P/CAF;
KW   transcription; histone acetyltransferase; HIV; infection; cancer;
KW   therapy; muscle differentiation.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Region          1763. .1966
FT                   /note= "P/CAF binding region"
PN   WO9803652-A2.
PD   29-JAN-1998.
PF   23-JUL-1997; U12877.
PR   23-JUL-1996; US-022273.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI   Howard BH, Nakatani Y;
DR   WPI; 98-120777/11.
DR   N-PSDB; V10092.
PT   New isolated p300/CBP-associated factor, P/CAF - used to develop
PT   products for modulating transcription, e.g. for treating HIV
PT   infection or cancers or for promoting muscle differentiation
PS   Disclosure; Page 76-81; 107pp; English.
CC   This polypeptide sequence comprises p300, a global transcriptional
CC   coactivator that is involved in the regulation of various
CC   DNA-binding transcriptional factors.  The invention relates to a
CC   novel human p300/CBP associated cofactor, P/CAF (see W40052), that
CC   modulates transcription through binding to p300 and CBP (see
CC   W40058).  The region (see W40055) of p300 that binds to P/CAF
CC   is claimed.  The invention provides methods of screening for
CC   compounds that inhibit or stimulate the transcription modulating
CC   and histone acetyltransferase activity of P/CAF and p300/CBP.
CC   Inhibitors can be used e.g. to inhibit HIV TAT-mediated
CC   transcription in the treatment of HIV infection.  Stimulators can
CC   be used e.g. to activate tumour suppressor p53 in the treatment of
CC   cancer or to activate the muscle differentiation factor MyoD to
CC   promote muscle differentiation.  The products can also be used to
CC   inhibit the cell cycle progression inducing effect of an
CC   oncoprotein which binds p300/CBP in a subject.  Also provided is
CC   a method for determining the amount of P/CAF in a sample by
CC   contacting the sample with the P/CAF binding region of p300 and
CC   determining the amount of P/CAF/p300 complex formed.
SQ   Sequence   2414 AA;

  Query Match             6.9%;  Score 110;  DB 1;  Length 2414;
  Best Local Similarity  34.0%;  Pred. No. 4.50e+00;
  Matches    17;  Conservative   13;  Mismatches  18;  Indels   2;  Gaps   2;

Db   2137 QAGVQRAGLPQQQPQQQLQPPMGGMSPQAQQ-MNMNHNTM-PSQFRDILR 2184
           : |||| |::   |::|| : ::||    || : :   :: |  ||| |:
Qy     65 RDGVQRLGVSANLPEEQLGALLAGMHTLLQQALRLPPTSLKPDTFRDQLQ 114


RESULT   11
ID   W19734 standard; Protein; 360 AA.
AC   W19734;
DT   18-SEP-1997  (first entry)
DE   Sugar biosynthesis enzyme EryCII.
KW   Polyketide; glycosylation; eryCII; D-desosamine; antimicrobial;
KW   antibiotic; antifungal; fungicide; anticancer; cytostatic;
KW   anthelmintic.
OS   Saccharopolyspora erythraea.
PN   WO9723630-A2.
PD   03-JUL-1997.
PF   23-DEC-1996; U20238.
PR   21-DEC-1995; US-576626.
PA   (ABBO ) ABBOTT LAB.
PI   Donadio S,  Katz L,  Staver MJ,  Summers RG;
DR   WPI; 97-351066/32.
DR   N-PSDB; T72684.
PT   New genes involved in sugar biosynthesis and attachment - used to
PT   generate polyketide antimicrobials etc. with altered pattern of
PT   glycosylation
PS   Disclosure; Fig 4A; 85pp; English.
CC   EryCII (W19734) is an enzyme involved in the biosynthesis of the
CC   sugar D-desosamine.  It is one of 10 enzymes (see also W19735-43)
CC   predicted to be involved in D-desosamine or L-mycarose biosynthesis
CC   and attachment that have been identified from gene clusters (T72684
CC   and T72685) of Saccharopolyspora erythraea.  Novel glycosylation-
CC   modified polyketides are produced by selectively altering,
CC   inactivating or augmenting the eryB and/or eryC genes encoding
CC   these sugar biosynthesis enzymes and introducing them into
CC   polyketide-producing microorganisms.
SQ   Sequence   360 AA;

  Query Match             6.7%;  Score 107;  DB 1;  Length 360;
  Best Local Similarity  25.2%;  Pred. No. 7.24e+00;
  Matches    33;  Conservative   36;  Mismatches  54;  Indels   8;  Gaps   7;

Db   150 MTAVLGAAGVLR--GAAWDARVS-LDAQLSPQQLAVTEAAVAALPADPALRAL-FAGAEM 205
          |:|| :|:  |:  | :  :||| | |||:|: :|     :: |      : | |  : :
Qy     1 MSAVGAATPYLHHPGDSHSGRVSFLGAQLPPE-VAAMARLLGDLDRSTFRKLLKFVVSSL 59

Db   206 TANTVVDAVLAVSAEPGLAE-RIADDPAAAQRTVAEVLRLHPA-LHLER-RTATAEVRLG 262
           ::   |:|  ::  : |:| :::   |: :  : : ||| |: |  :  |    |: :
Qy    60 QGEDCRDGVQRLGVSANLPEEQLGALLAGMHTLLQQALRLPPTSLKPDTFRDQLQELCIP 119

Db   263 EHVIGEGEEVV 273
          : ::|:   ||
Qy   120 QDLVGDLASVV 130


RESULT   12
ID   W99387 standard; Protein; 361 AA.
AC   W99387;
DT   08-JUN-1999  (first entry)
DE   S.erythraea dTDP-4-keto-L-6-deoxyhexose-3,4-isomerase.
KW   Gene cluster; bacterium; enzyme; macrolide; antibiotic; erythromycin;
KW   secondary metabolite; eryBII; eryCIII; eryCII; hybridisation; probe;
KW   glycosylation; macrolactone; oleandomycin.
OS   Saccharopolyspora erythraea.
PN   WO9905283-A2.
PD   04-FEB-1999.
PF   21-JUL-1998; F01593.
PR   12-JUN-1998; FR-007411.
PR   25-JUL-1997; FR-009458.
PA   (HMRI ) HOECHST MARION ROUSSEL.
PI   Cortes J, Gaisser S, Leadlay P, Michel JM, Raynal MC,
PI   Salah-Bey K, Fromentin C, Mendez C, Salas JA;
DR   WPI; 99-142938/12.
DR   N-PSDB; X25772.
PT   New nucleic acid sequences encoding enzymes involved in macrolide
PT   biosynthesis - useful for producing hybrid secondary metabolites,
PT   particularly erythromycin analogues
PS   Claim 6; Fig 2; 221pp; French.
CC   This sequence represents the dTDP-4-keto-L-6-deoxyhexose-3,4-isomerase
CC   enzyme encoded by the eryCII gene from the eryG-eryAIII gene cluster from
CC   the Gram-positive bacterium Saccharopolyspora erythraea.  This gene
CC   cluster encodes enzymes involved in the production of the macrolide
CC   antibiotic erythromycin as a secondary metabolite.  The erythromycin gene
```

```
CC   cluster spans approximately 53 kb and contains at least 20 open reading
CC   frames (ORF).  The genes are used to produce hybrid secondary metabolites
CC   in S.erythraea, i.e. erythromycin analogues which may have improved
CC   properties or as hybridisation probes for isolating homologous genes
CC   involved in glycosylation of macrolactones in macrolide-producing
CC   strains (specifically oleandomycin-producing strains of Streptomyces
CC   antibioticus).
SQ   Sequence   361 AA;

  Query Match             6.7%;  Score 107;  DB 1;  Length 361;
  Best Local Similarity  25.2%;  Pred. No. 7.24e+00;
  Matches    33;  Conservative   36;  Mismatches  54;  Indels   8;  Gaps   7;

Db    151 MTAVLGAAGVLR--GAAWDARVS-LDAQLSPQQLAVTEAAVAALPADPALRAL-FAGAEM 206
          |:|| :|:  |:  | :  :||| | |||:|: :|     :: |      : | |  : :
Qy      1 MSAVGAATPYLHHPGDSHSGRVSFLGAQLPPE-VAAMARLLGDLDRSTFRKLLKFVVSSL 59

Db    207 TANTVVDAVLAVSAEPGLAE-RIADDPAAAQRTVAEVLRLHPA-LHLER-RTATAEVRLG 263
          ::   |:|  ::  : |:| :::   |: :  : : ||| |: |  :  |    |: :
Qy     60 QGEDCRDGVQRLGVSANLPEEQLGALLAGMHTLLQQALRLPPTSLKPDTFRDQLQELCIP 119

Db    264 EHVIGEGEEVV 274
          : ::|:   ||
Qy    120 QDLVGDLASVV 130


RESULT   13
ID   R77317 standard; Protein; 339 AA.
AC   R77317;
DT   07-FEB-1996  (first entry)
DE   Protein activated lipase B.
KW   Lipase B; activate; glutathionine transferase.
OS   Pseudomonas aeruginosa.
PN   J07135971-A.
PD   30-MAY-1995.
PF   05-JUL-1993; 165561.
PR   05-JUL-1993; JP-165561.
PA   (TOYM ) TOYOBO KK.
DR   WPI; 95-227395/30.
DR   N-PSDB; Q91399.
PT   A new protein for activating lipase - may be produced as a fusion
PT   protein with glutathione transferase, used in the production of
PT   lipase
PS   Claim 1; Page 6-7; 9pp; Japanese.
CC   R77317 represents a new protein-activated lipase. The protein is
CC   produced as a fusion protein with glutathionine transferase. The
CC   protein has an amino acid sequence of 339 residues and a molecular
CC   weight of 39 kD. The method used to produce the new protein yields
CC   active lipase of very high purity without using any complex
CC   purification process such as chromatography.
SQ   Sequence   339 AA;

  Query Match             6.6%;  Score 106;  DB 1;  Length 339;
  Best Local Similarity  30.5%;  Pred. No. 8.48e+00;
  Matches    18;  Conservative   21;  Mismatches  17;  Indels   3;  Gaps   2;

Db    207 LSTEEKAAAIDRL--RASLPEDQQESVLPQLQSELQQQTAALQAAGAGPEAIRQMRQQL 263
          |  |:   :::||   |:|||:|  ::|: ::: | ||:  | :::  |:::|:  |:|
Qy     59 LQGEDCRDGVQRLGVSANLPEEQLGALLAGMHT-LLQQALRLPPTSLKPDTFRDQLQEL 116


RESULT   14
ID   W99798 standard; Protein; 727 AA.
AC   W99798;
DT   16-JUN-1999  (first entry)
DE   Human VRRP-1 (VR2) capsaicin receptor.
KW   VR1; capsaicin receptor; VR2; VRRP-1; analgesic; diagnosis;
KW   human disease; painful syndrome.
OS   Homo sapiens.
PN   WO9909140-A1.
PD   25-FEB-1999.
PF   20-AUG-1998; U17466.
PR   22-JAN-1998; US-072151.
PR   20-AUG-1997; US-915461.
PA   (REGC ) UNIV CALIFORNIA.
PI   Brake A, Caterina M, Julius DJ;
DR   WPI; 99-181023/15.
PT   New capsaicin receptor polypeptide - useful for screening or
PT   characterising capsaicin receptor-binding compounds
PS   Claim 4; Page 86-88; 99pp; English.
CC   The present sequence is an isolated capsaicin receptor polypeptide
CC   (I). Capsaicin polypeptides are useful for identifying binding compounds
CC   which affect cellular responses. Preferably this is for identifying a
CC   compound that binds (I) and affects a cellular response associated with
CC   capsaicin biological activity (e.g. intracellular calcium flux). The
CC   polypeptides and host cells are useful for detecting a vanilloid
CC   compound (an essential structural component of capsaicin) from natural
CC   products by detecting an alteration of intracellular response associated
CC   with capsaicin receptor activity, preferably an alteration of
CC   intracellular calcium levels, and are useful for screening for compounds
CC   for use in analgesics. Capsaicin receptor polypeptides and antibodies
CC   are useful for diagnosis and treatment of human diseases and painful
CC   syndromes. The transgenic mammals can be used to screen for capsaicin
CC   receptor antagonists and agonists. Prior art methods for screening or
CC   characterising new capsaicin receptor-binding compounds relied on assays
CC   using sensory neurons in culture or in intact animals. The new
CC   polypeptides provide a more sensitive screen.
SQ   Sequence   727 AA;

  Query Match             6.6%;  Score 106;  DB 1;  Length 727;
  Best Local Similarity  29.3%;  Pred. No. 8.48e+00;
  Matches    24;  Conservative   24;  Mismatches  26;  Indels   8;  Gaps   8;

Db    236 ASLQATDSQGNTVLHALVMISDNSAENI-ALVTSMYDGLLQAGARLCPT-VQLEDIRN-L 292
          :|||: | : : | : | : :: : | : ||:::|   ||| : || || :  : :|: |
Qy     57 SSLQGEDCR-DGV-QRLGVSANLPEEQLGALLAGMHT-LLQQALRLPPTSLKPDTFRDQL 113

Db    293 QDL-TPLKLAAKEGKIXIFXRH 313
          |:|  |  | :  : : :|  :
Qy    114 QELCIPQDLVGDLASV-VFGSQ 134


RESULT   15
ID   R47213 standard; Protein; 344 AA.
AC   R47213;
DT   27-JUL-1994  (first entry)
DE   Lipase modulator.
KW   Pseudomonas; cloning; lipase gene; lipase modulator gene; lipids;
KW   oils; fats; detergents; diagnostic reagents.
OS   Pseudimonas pseudoalcaligenes.
PN   WO9402617-A.
PD   03-FEB-1994.
PF   23-JUL-1993; E01995.
PR   23-JUL-1992; EP-202281.
PA   (KONN ) GIST-BROCADES NV.
PI   Cox MMJ,  Gerritse G,  Quax WJ;
DR   WPI; 94-048875/06.
DR   N-PSDB; Q56106.
PT   Increasing lipase prodn. in Pseudomonas species - by cloning a
PT   lipase gene and a lipase modulator gene into a homologous
PT   Pseudomonas species
PS   Disclosure; Page 23; 44pp; English.
CC   The lipase gene from Pseudomonas pseudoalcaligenes M1 was cloned in
CC   E. coli as in EP-334462.  The DNA sequence of insert PvuI/EcoRI
CC   contains two open reading frames, one encoding the lipase gene, the
CC   other encoding a putative lipase modulator gene which increases the
CC   lipase productivity in a homologous host cell.  The lipase obtd. is
CC   used for hydrolysing lipids in e.g. fats and oil processing,
CC   detergents or diagnostic reagents.
CC   See also R47212.
SQ   Sequence   344 AA;

  Query Match             6.5%;  Score 104;  DB 1;  Length 344;
  Best Local Similarity  21.7%;  Pred. No. 1.16e+01;
  Matches    26;  Conservative   38;  Mismatches  48;  Indels   8;  Gaps   5;
```

```
Db    199 LQRLAIRHDQTLDDQQKAEALDRLRASLPEELQALLAPQLQAE-LRQQTAALQAQGASAA 257
          :|||::  :  | ::| :  |  ::: | : |: | :  | :: :|:|   ||    :
Qy     68 VQRLGVSAN--LPEEQLGALLAGMHTLLQQALR-LPPTSLKPDTFRDQ---LQELCIPQD 121

Db    258 QIQQLRLQLVGAEATARLEALDQQRQQWRQRLADYRREKARVLANDGLSESDKQAAIAEL 317
           : :|   : |::    |::: ||:  |  ::||:|:      ::  :|: |   : : :|
Qy    122 LVGDLASVVFGSQRPL-LDSVAQQQGAWLPHVADFRWRVDVAISTSALARSLQPSVLMQL 180
```

Search completed: Fri May 12 10:36:01 2000
Job time : 34 secs.

****

MPSRCH (TM)

****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 12 10:24:14 2000; MasPar time 81.94 Seconds
189.543 Million cell updates/sec

Tabular output not generated.

Title: >US-09-223-796-2
Description: (1-224) from US09223796.pep
Perfect Score: 1552
Sequence: 1 MSALGAAAPYLHHPADSHSG.....ALVLKEMAELEKKCERKLQD 224

Scoring table: PAM 150
Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl12
1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 45.804; Variance 101.760; scale 0.450

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	124	8.0	195	11	Q63829	BUP=5'OF BMI-1 PROVIRA	4.75e-03
2	121	7.8	491	2	P73944	HYPOTHETICAL 55.6 KD P	1.15e-02
3	119	7.7	202	4	Q9Y6G5	PTD002.	2.05e-02
4	120	7.7	649	4	O43642	HEAT SHOCK PROTEIN 75.	1.54e-02
5	118	7.6	512	10	Q9ZW95	F5A8.3 PROTEIN.	2.74e-02
6	118	7.6	676	5	Q27512	NEX-2 PROTEIN.	2.74e-02
7	110	7.1	706	2	P72747	HYPOTHETICAL 79.8 KD P	2.64e-01
8	107	6.9	470	5	O62159	F13G11.1 PROTEIN.	6.02e-01
9	107	6.9	913	5	P90742	CODED FOR BY C. ELEGAN	6.02e-01
10	105	6.8	279	2	Q9Z5S3	FLAGELLIN.	1.03e+00
11	104	6.7	226	14	Q98184	MC013L.	1.35e+00
12	104	6.7	228	1	Q9YBI8	228AA LONG HYPOTHETICA	1.35e+00
13	104	6.7	302	9	Q37840	P2 J HOMOLOG.	1.35e+00
14	104	6.7	382	2	Q9X7Y2	PROBABLE ACYL-COA DEHY	1.35e+00
15	104	6.7	603	11	O70211	INSULIN-LIKE GROWTH FA	1.35e+00
16	104	6.7	1116	2	P73507	DNA POLYMERASE III SUB	1.35e+00
17	104	6.7	4845	11	O88738	UBIQUITIN-CONJUGATING	1.35e+00
18	102	6.6	164	2	O87525	GALACTOSE-1-PHOSPHATE	2.30e+00
19	102	6.6	451	2	P72863	HYPOTHETICAL 49.7 KD P	2.30e+00
20	103	6.6	555	2	Q9Z924	FRUCTOSE-6-P PHOSPHOTR	1.77e+00
21	103	6.6	902	4	O43632	SPINDLE POLE BODY PROT	1.77e+00
22	103	6.6	1101	10	Q9ZQT7	L-ZIP+NBS+LRR.	1.77e+00
23	100	6.4	228	2	Q9ZHD3	PUTATIVE TRANSCRIPTION	3.90e+00
24	99	6.4	260	13	Q9W644	PROLIFERATING CELL NUC	5.05e+00
25	99	6.4	266	2	Q47220	FIMBRIAL PROTEIN.	5.05e+00
26	100	6.4	362	1	Q9YF28	362AA LONG HYPOTHETICA	3.90e+00
27	99	6.4	371	2	P74657	UDP-N-ACETYLGLUCOSAMIN	5.05e+00
28	99	6.4	933	3	O14310	NUCLEOPORIN.	5.05e+00
29	99	6.4	1032	3	O94324	SERINE-THREONINE PROTE	5.05e+00
30	99	6.4	1547	5	Q26471	BLACKJACK.	5.05e+00
31	98	6.3	229	2	P96415	30S RIBOSOMAL PROTEIN	6.54e+00
32	98	6.3	295	5	Q20233	F40F8.7 PROTEIN.	6.54e+00
33	98	6.3	340	2	O84915	GALR-LACI HOMOLOG.	6.54e+00
34	98	6.3	344	2	O53595	HYPOTHETICAL 37.0 KD P	6.54e+00
35	96	6.2	312	2	P72131	PTXR.	1.09e+01
36	96	6.2	312	2	O30374	TRANSCRIPTIONAL ACTIVA	1.09e+01
37	97	6.2	386	2	Q9X8A4	HYPOTHETICAL 40.9 KD P	8.45e+00
38	97	6.2	669	2	Q55421	ELONGATION FACTOR EF-G	8.45e+00
39	96	6.2	960	10	O82318	RECEPTOR-LIKE PROTEIN	1.09e+01
40	96	6.2	1329	5	O76356	C45G7.6 PROTEIN.	1.09e+01
41	96	6.2	1846	5	O61776	F56A6.2 PROTEIN.	1.09e+01
42	96	6.2	2228	10	O48579	HUMAN MI-2 AUTOANTIGEN	1.09e+01
43	95	6.1	287	2	O53288	ABC TRANSPORTER ATP-BI	1.40e+01
44	95	6.1	315	2	O83061	CONSERVED HYPOTHETICAL	1.40e+01
45	95	6.1	452	10	O49368	HYPOTHETICAL 50.0 KD P	1.40e+01

ALIGNMENTS

RESULT 1
ID Q63829 PRELIMINARY; PRT; 195 AA.
AC Q63829;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE BUP=5'OF BMI-1 PROVIRAL INSERTION LOCUS.
GN BUP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93165034.
RA HAUPT Y., BARRI G., ADAMS J.M.;
RT "Nucleotide sequence of bup, an upstream gene in the bmi-1 proviral
RT insertion locus.";
RL Mol. Biol. Rep. 17:17-20(1992).
DR EMBL; S54914; AAB25378.1; -.
DR MGD; MGI:88218; Bup.
SQ SEQUENCE 195 AA; 22037 MW; 827A84F1 CRC32;

Query Match 8.0%; Score 124; DB 11; Length 195;
Best Local Similarity 22.1%; Pred. No. 4.75e-03;
Matches 21; Conservative 33; Mismatches 37; Indels 4; Gaps 4;

```
Db      98 ELFCTEYQNNKNSLETLLGSIGRSLPHITDVSWRLEYQIKTNQLHKMYRPGYLVTLNVEN 157
           :|    : : :  |:::    | ||||::   ||::  | |:   :  :|: |: |:: :
Qy     125 DLASLAFGSQRPLLDSVAQQQGSSLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQLKLTD 184

Db     158 NDSQSYPEINFSCNMEQLQDLVG-KLKDASKSLER 191
             :: : |: :: : ::|:  |:  ||: :  ||:
Qy     185 GSAHRF-EVPIA-KFQELRYSVALVLKEMAE-LEK 216
```

RESULT 2
ID P73944 PRELIMINARY; PRT; 491 AA.
AC P73944;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE HYPOTHETICAL 55.6 KD PROTEIN.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PCC6803;
RA   TABATA S.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PCC6803;
RX   MEDLINE; 97061201.
RA   KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA   MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA   HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
RA   SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
RA   TABATA S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
RL   DNA Res. 3:109-136(1996).
DR   EMBL; D90911; BAA18010.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   491 AA;  55645 MW;  6EFDFCFA CRC32;

  Query Match             7.8%;  Score 121;  DB 2;  Length 491;
  Best Local Similarity  23.1%;  Pred. No. 1.15e-02;
  Matches    27;  Conservative   36;  Mismatches  47;  Indels   7;  Gaps   7;

Db    121 RELNRN-YTKALQLLRRVSEQSRNLRQEVANLSAERAQLNEQK-DSLLAESSELQSQVKL 178
          ::|:|: : | | |:   : :::: |: |  |:|  | |:|::   ||| :  |  |  |
Qy     41 KDLDRSTFRKLLKLVV-GALHGKDCREAVEQLGAS-ANLSEERLAVLLAGTHTLLQQA-L 97

Db    179 RDQELSKRQERIAQQEKVLARQREQVQSLEKRFASLEAQRQQLQAEINQRDTKIDQL 235
          |    | : : : ::   |:  :: :  |   :| : :||  |::   |: : : ::
Qy     98 RLPPASLKPDAFQEELQELGIPQDLIGDLAS-LA-FGSQRPLLDSVAQQQGSSLPHV 152


RESULT    3
ID   Q9Y6G5      PRELIMINARY;      PRT;   202 AA.
AC   Q9Y6G5;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   PTD002.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PITUITARY TUMOR;
RA   FU G., HUANG Q., SONG H., PENG J., ZHANG Q., MAO M., DAI M., MAO Y.,
RA   ZHOU J., CHEN Z., CHEN J.;
RT   "Human PTD002 mRNA, complete cds.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF078857; AAD44489.1; -.
SQ   SEQUENCE   202 AA;  22966 MW;  51317189 CRC32;

  Query Match             7.7%;  Score 119;  DB 4;  Length 202;
  Best Local Similarity  22.0%;  Pred. No. 2.05e-02;
  Matches    38;  Conservative   53;  Mismatches  72;  Indels  10;  Gaps   8;

Db     12 SPSMKKAVSLINAIDTGRFPRLLTRILQKLHLKAESSFSEEEEEKLQAAFSLEKQDLHLV 71
          || :     |:: :| : | :||   ::  || | :     |  |: | |: :|  : | ::
Qy     30 SPEVTAVAQLLKDLDRSTFRKLLKLVVGALHGK-DCR--EAVEQ-LGASANLSEERLAVL 85

Db     72 LETISFILEQAV-YH--NVKPAALQQQLENIHLRQDKAEAFVN-TWSSMGQETVEKFRQR 127
          |     :|:||:     ::|| |:|::|::: : ||    : : : :|  :  ::   |:
Qy     86 LAGTHTLLQQALRLPPASLKPDAFQEELQELGIPQDLIGDLASLAFGSQ-RPLLDSVAQQ 144

Db    128 ILAP-CKLETVGWQLNLQMAHSAQAKLKSPQAVLQLGVNNEDSKSLEKVLVEF 179
            ::   :    |:::: :: |||::   |  ::|| : :  :  :|  :  |
Qy    145 QGSSLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQLKLTDGSAHRFEVPIAKF 197


RESULT    4
ID   O43642      PRELIMINARY;      PRT;   649 AA.
AC   O43642;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   HEAT SHOCK PROTEIN 75.
GN   HSP75.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96347537.
RA   CHEN C.F., CHEN Y., DAI K., CHEN P.L., RILEY D.J., LEE W.H.;
RT   "A new member of the hsp90 family of molecular chaperones interacts
RT   with the retinoblastoma protein during mitosis and after heat shock.";
RL   Mol. Cell. Biol. 16:4691-4699(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   CHEN C.F., CHEN Y., DAI K., CHEN P.L., RILEY D.J., LEE W.H.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF043254; AAC02679.1; -.
DR   HSSP; P02829; 1AH8.
DR   PFAM; PF00183; HSP90; 2.
DR   PRINTS; PR00775; HEATSHOCK90.
KW   Heat shock.
SQ   SEQUENCE   649 AA;  74017 MW;  CCAF787C CRC32;

  Query Match             7.7%;  Score 120;  DB 4;  Length 649;
  Best Local Similarity  22.3%;  Pred. No. 1.54e-02;
  Matches    23;  Conservative   40;  Mismatches  34;  Indels   6;  Gaps   6;

Db     43 KKLLDIVARSLYS-E-KEVFIRELISNASDALEKLRHKLVSDGQALPEMEIHLQTNAEKG 100
          :||| :|  :| : : :|  : :| ::|: : |:|   |::  ::| :  ::|   : |
Qy     49 RKLLKLVVGALHGKDCREA-VEQLGASANLSEERLAV-LLAGTHTLLQQALRLPPASLKP 106

Db    101 TITIQD-TGIGMTQEELVSNLGTIARSGSKAFLDALQNQAEAS 142
             ::   :|: |: |:::|:::| :: :::||::  |  :|
Qy    107 DAFQEELQELGIPQD-LIGDLASLAFGSQRPLLDSVAQQQGSS 148


RESULT    5
ID   Q9ZW95      PRELIMINARY;      PRT;   512 AA.
AC   Q9ZW95;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   F5A8.3 PROTEIN.
GN   F5A8.3.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC   Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   FEDERSPIEL N.A., PALM C.J., CONWAY A.B., KURTZ D.B., CONWAY A.R.,
RA   AU M., ARAUJO R., BUEHLER E., DEWAR K., FENG J., KIM C., LI Y.,
RA   OJI O., OSBORNE B.I., SHINN P., SUN H., TORIUMI M., VYSOTSKAIA V.S.,
RA   YU G., ECKER J., THEOLOGIS A., DAVIS R.W.;
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC004146; AAD10659.1; -.
SQ   SEQUENCE   512 AA;  58463 MW;  9C771B39 CRC32;

  Query Match             7.6%;  Score 118;  DB 10;  Length 512;
  Best Local Similarity  32.1%;  Pred. No. 2.74e-02;
  Matches    25;  Conservative   20;  Mismatches  28;  Indels   5;  Gaps   5;

Db    333 WTLMLLAHNPTWQDNVRDEVRQVCGQDGVPSVEQLSSL-TSLNKVINESLRLYPPATLLP 391
          : |: |   :    : |: | |: |  :  | |:|: | :: : :: ::||| |||:| |
Qy     49 RKLLKLVVGALHGKDCREAVEQL-GASANLSEERLAVLLAGTHTLLQQALRL-PPASLKP 106
```

```
Db    392 RMAFEDIKLGDLIIPKGL 409
            ||::  | :| ||  |
Qy    107 D-AFQE-ELQELGIPQDL 122


RESULT    6
ID   Q27512      PRELIMINARY;      PRT;   676 AA.
AC   Q27512;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   NEX-2 PROTEIN.
GN   NEX-2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BERKS M.;
RL   Submitted (JAN-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
CC   -!- FUNCTION: CALCIUM/PHOSPHOLIPID-BINDING PROTEIN WHICH PROMOTES
CC       MEMBRANE FUSION AND IS INVOLVED IN EXOCYTOSIS.
CC   -!- SIMILARITY: TO OTHER PROTEINS OF THE ANNEXIN FAMILY.
DR   EMBL; Z29443; CAA82571.1; -.
DR   HSSP; P79134; 1AVC.
DR   PROSITE; PS00223; ANNEXIN; 1.
DR   PFAM; PF00191; annexin; 4.
DR   PRINTS; PR00196; ANNEXIN.
KW   Annexin; Calcium/phospholipid-binding; Repeat.
SQ   SEQUENCE   676 AA;  74969 MW;  00E36C79 CRC32;

  Query Match             7.6%;  Score 118;  DB 5;  Length 676;
  Best Local Similarity  25.7%;  Pred. No. 2.74e-02;
  Matches    37;  Conservative   31;  Mismatches  67;  Indels   9;  Gaps   8;

Db     59 C-ELVKESRSHEMLSEEQLATLYDCATTSTRLPNDFADDQEDLKNIIRYGELFKACHAIN 117
          | | | :  :   ||||:|| |   | | | |   :      ||    : | :    :|
Qy     64 CREAVEQLGASANLSEERLAVLL--AGTHTLLQQALRLPPASLKPDA-FQEELQE-LGI- 118

Db    118 STDFIQKSEDLK-DEEKVALERIVEQKLAESAKNEKDIAWNVNIVVANSYVAKSLRPVIN 176
          : |:|     |    ::  |: : :|  : |  :   : | |:: :: |  ::||:| :
Qy    119 PQDLIGDLASLAFGSQRPLLDSVAQQQ-GSSLPHVSYFRWRVDVAISTSAQSRSLQPSVL 177

Db    177 IQMPTVGGDTN-FEFDIDSFAQFR 199
          :|:    | :: ||  |  | ::|
Qy    178 MQLKLTDGSAHRFEVPIAKFQELR 201


RESULT    7
ID   P72747      PRELIMINARY;      PRT;   706 AA.
AC   P72747;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   HYPOTHETICAL 79.8 KD PROTEIN.
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PCC6803;
RA   TABATA S.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PCC6803;
RX   MEDLINE; 97061201.
RA   KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA   MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA   HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
RA   SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
RA   TABATA S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
RL   DNA Res. 3:109-136(1996).
DR   EMBL; D90900; BAA16762.1; -.
DR   PFAM; PF00563; DUF2; 1.
DR   PFAM; PF00990; DUF9; 1.
DR   PFAM; PF00498; FHA; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   706 AA;  79831 MW;  706E0D1E CRC32;

  Query Match             7.1%;  Score 110;  DB 2;  Length 706;
  Best Local Similarity  27.7%;  Pred. No. 2.64e-01;
  Matches    23;  Conservative   29;  Mismatches  26;  Indels   5;  Gaps   5;

Db    585 RRAIEQLATIGVNLSLDDFGIGL-GTLSCLQQ-FKIPAVKIH-ESFIKDLEQSPVNEAII 641
          | |:|||:: : ||| : ::: | || : ||| :::|:  :  ::|  :|::  : : :|
Qy     65 REAVEQLGA-SANLSEERLAVLLAGTHTLLQQALRLPPASLKPDAFQEELQELGIPQDLI 123

Db    642 TSIMTLGRKLGVRIISEGVETQQ 664
            : :|:   : | : ::|  ||
Qy    124 GDLASLAFG-SQRPLLDSVAQQQ 145


RESULT    8
ID   O62159      PRELIMINARY;      PRT;   470 AA.
AC   O62159;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   F13G11.1 PROTEIN.
GN   F13G11.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   AINSCOUGH R.;
RL   Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z83317; CAB05899.1; -.
DR   PFAM; PF00751; DM-domain; 1.
SQ   SEQUENCE   470 AA;  51820 MW;  CC968597 CRC32;
```

```
  Query Match           6.9%;  Score 107;  DB 5;  Length 470;
  Best Local Similarity  26.8%;  Pred. No. 6.02e-01;
  Matches   30;  Conservative   30;  Mismatches  48;  Indels   4;  Gaps   4;

Db    358 ALSNLSAGQELSSFILHSVVLQAQQAQQKTAQQLIPTSM-PSELESTLSLLRLQQSSLGD 416
          |:  |:|: :||   | :|:| : ::    | :| |:|: |  ::  |  | : |  :||
Qy     67 AVEQLGASANLSEERL-AVLLAGTHTLLQQALRLPPASLKPDAFQEELQELGIPQDLIGD 125

Db    417 LIKLS-GIQLPTLPTTQPMLTPQMPTLPMMQKALPTVLPGSFQP-FLQPTLF 466
          |  |: | | | | :      : :| ::  :  :   :: | |:  |||:::
Qy    126 LASLAFGSQRPLLDSVAQQQGSSLPHVSYFRWRVDVAISTSAQSRSLQPSVL 177


RESULT    9
ID   P90742      PRELIMINARY;      PRT;   913 AA.
AC   P90742;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   CODED FOR BY C. ELEGANS CDNA YK40D1.5.
GN   C09B9.7.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J.) LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   BRADSHAW H.;
RL   Submitted (MAR-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   WATERSTON R.;
RL   Submitted (FEB-1996) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   WATERSTON R.;
RL   Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U50069; AAB37561.1; -.
SQ   SEQUENCE   913 AA;  103125 MW;  2D649BD4 CRC32;

  Query Match           6.9%;  Score 107;  DB 5;  Length 913;
  Best Local Similarity  23.2%;  Pred. No. 6.02e-01;
  Matches   26;  Conservative   36;  Mismatches  45;  Indels   5;  Gaps   5;

Db    211 IGLTTDEAMSDVNQILELVNSVESDQFKKHSIGFFRGILSSKHLERESRDPWLGKVLNVS 270
          :::   :: ::|  : :|:: ::   |:|  : :  | | :|   ||: :  ||   |:|
Qy     22 VSFLGSQPSPEVTAVAQLLKDLDRSTFRKL-LKLVVGALHGKDC-REAVEQ-LGASANLS 78

Db    271 EEHLASITSGLQPLIKASDHLETVNDKLMSVRPNALSRSLGYFQSFVTTMSS 322
          ||:|| : :| : |:  : :|   : |  : : : | : ||  | ::  ::|
Qy     79 EERLAVLLAGTHTLLQQALRLPPASLKPDAFQ-EEL-QELGIPQDLIGDLAS 128


RESULT   10
```

```
ID   Q9Z5S3      PRELIMINARY;      PRT;   279 AA.
AC   Q9Z5S3;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   FLAGELLIN.
GN   FLA.
OS   Zymomonas mobilis.
OC   Bacteria; Proteobacteria; alpha subdivision; Zymomonas group;
OC   Zymomonas.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ZM4;
RA   UM H.W., KANG H.S.;
RT   "Sequence analysis of 41A4 fosmid clone of Zymomonas mobilis.";
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF124349; AAD19738.1; -.
SQ   SEQUENCE   279 AA;  29162 MW;  A449F61B CRC32;

  Query Match           6.8%;  Score 105;  DB 2;  Length 279;
  Best Local Similarity  23.0%;  Pred. No. 1.03e+00;
  Matches   26;  Conservative   36;  Mismatches  44;  Indels   7;  Gaps   6;

Db    150 QTGTTSAAATDRVTISISGLSTL--KTI-MSNATLKVATSTDALKALSTINKALTGINTT 206
          | |:::  : :|::: ::|  ||   :: :: |:||  :  : |  |: |   | | : :
Qy     70 QLGASANLSEERLAVLLAGTHTLLQQALRLPPASLKPDAFQEELQELG-IPQDLIG-DLA 127

Db    207 RAGLGASESRLSSVVNNLTTNVTNLSDARSRIQ-DTDYSSETTNLARSQILSQ 258
            ::|:  : | ||     ::: ::|  | |::     |::: :|  | :| |
Qy    128 SLAFGSQRPLLDSVAQQQGSSLPHVSYFRWRVDVAISTSAQSRSLQPS-VLMQ 179


RESULT   11
ID   Q98184      PRELIMINARY;      PRT;   226 AA.
AC   Q98184; O12880; O12599;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   MC013L.
GN   MC013L OR B-M,N,L.3.
OS   Molluscum contagiosum virus subtype 1 (MCVI).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Molluscipoxvirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96325459.
RA   SENKEVICH T.G., BUGERT J.J., SISLER J.R., KOONIN E.V., DARAI G.,
RA   MOSS B.;
RT   "Genome sequence of a human tumorigenic poxvirus: prediction of
RT   specific host response-evasion genes.";
RL   Science 273:813-816(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   SENKEVICH T.G., BUGERT J.J., SISLER J.R., KOONIN E.V., DARAI G.,
RA   MOSS B.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97093414.
RA   MARTIN-GALLARDO A., MORATILLA M., FUNES J.M., AGROMAYOR M., NUNEZ A.,
RA   VARAS A.J., COLLADO M., VALENCIA A., LOPEZ-ESTEBARANZ J.L.,
RA   ESTEBAN M.;
RT   "Sequence analysis of a Molluscum contagiosum virus DNA region which
RT   includes the gene encoding protein kinase 2 and other genes with
RT   unique organization.";
RL   Virus Genes 13:19-29(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RA   MORATILLA M., AGROMAYOR M., NUNEZ A., FUNES J.M., VARAS A.J.,
RA   LOPEZ-ESTEBARANZ J.L., ESTEBAN M., MARTIN-GALLARDO A.;
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
```

```
RA   MORATILLA M., AGROMAYOR M., NUNEZ A., FUNES J.M., VARAS A.J.,
RA   LOPEZ-ESTEBARANZ J.L., COLLADO M., PAVON M., MARTIN-GALLARDO A.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U60315; AAC55141.1; -.
DR   EMBL; U86894; AAB57933.1; -.
DR   HSSP; P25685; 1HDJ.
DR   PFAM; PF00226; DnaJ; 1.
SQ   SEQUENCE   226 AA;  25301 MW;  584D5F47 CRC32;

  Query Match             6.7%;  Score 104;  DB 14;  Length 226;
  Best Local Similarity  30.3%;  Pred. No. 1.35e+00;
  Matches    27;  Conservative   23;  Mismatches  34;  Indels   5;  Gaps   5;

Db    132 EAVEGLERVLHLPGGRLLTVELAPFQALETRSLRFPGYGLHYSGMAGELRVALRVRVPDA 191
          |||| |    :|:  || :| ||  ::|  ::||:|  :|   :   ||:  |   :|:
Qy     66 EAVEQLGASANLSEERL-AVLLAGTHTLLQQALRLPPASLKPDAFQEELQ-EL-G-IPQD 121

Db    192 LRAQPAQL-LSALRTALAGANMHERASVP 219
          | :: | | ::: |  | :   :: :|:|
Qy    122 LIGDLASLAFGSQRPLLDSVAQQQGSSLP 150


RESULT   12
ID   Q9YBI8      PRELIMINARY;      PRT;   228 AA.
AC   Q9YBI8;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   228AA LONG HYPOTHETICAL PROTEIN.
GN   APE1610.
OS   Aeropyrum pernix.
OC   Archaea; Crenarchaeota; Aeropyrum.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K1;
RX   MEDLINE; 99310339.
RA   KAWARABAYASI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,
RA   JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,
RA   HOSOYAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,
RA   TAKAMIYA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,
RA   YAMAZAKI J., KUSHIDA N., OGUCHI A., AOKI K., KUBOTA K., NAKAMURA Y.,
RA   NOMURA N., SAKO Y., KIKUCHI H.;
RT   "Complete genome sequence of an aerobic hyper-thermophilic
RT   crenarchaeon, Aeropyrum pernix K1.";
RL   DNA Res. 6:83-101(1999).
DR   EMBL; AP000062; BAA80610.1; -.
SQ   SEQUENCE   228 AA;  25890 MW;  A421A9B2 CRC32;

  Query Match             6.7%;  Score 104;  DB 1;  Length 228;
  Best Local Similarity  32.2%;  Pred. No. 1.35e+00;
  Matches    29;  Conservative   22;  Mismatches  33;  Indels   6;  Gaps   6;

Db    131 VVQLVEKDVKQSPLRRKLQRLKLPDLAAEAYATEGDALAVARELWG-RLARLGAVFRVSI 189
          | ||: ||: :| :| || :| :  | :       : |: : :|   ||| | |  : ::
Qy     36 VAQLL-KDLDRSTFR-KLLKLVVGALHGKDCREAVEQLGASANLSEERLAVLLAGTH-TL 92

Db    190 VPSRRLVLAVASPRPDSFGPVVEEVLSIVQ 219
          :  : | |: || :||:|   ::| |:| |
Qy     93 LQ-QALRLPPASLKPDAFQEELQE-LGIPQ 120


RESULT   13
ID   Q37840      PRELIMINARY;      PRT;   302 AA.
AC   Q37840;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   P2 J HOMOLOG.
GN   L.
OS   Bacteriophage 186.
OC   Viruses; dsDNA viruses, no RNA stage; Tailed phages; Myoviridae.
RN   [1]
RP   SEQUENCE OF 241-302 FROM N.A.
RA   XUE Q.;
RT   "Studies on the tail region of the temperate coliphage 186 genome.";
RL   Thesis (1993), University of Adelaide, Australia.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98371265.
RA   PORTELLI R., DODD I.B., XUE Q., EGAN J.B.;
RT   "The late-expressed region of the temperate coliphage 186 genome.";
RL   Virology 248:117-130(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   DODD I.B., EGAN J.B.;
RT   "Bacteriophage 186 complete genome.";
RL   Submitted (OCT-1995) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   XUE Q., EGAN J.B.;
RL   Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   XUE Q., EGAN J.B.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U32222; AAC34162.1; -.
FT   VARIANT      33     33       S -> *.
SQ   SEQUENCE   302 AA;  33005 MW;  5A210700 CRC32;

  Query Match             6.7%;  Score 104;  DB 9;  Length 302;
  Best Local Similarity  25.2%;  Pred. No. 1.35e+00;
  Matches    37;  Conservative   37;  Mismatches  67;  Indels   6;  Gaps   6;

Db    135 MSVAGSTGAYEFH-GRSADGRVADISVISPSPACVTISVLSRENNGAASDELQSIVRNAL 193
          ||  |::::|  | : |  |||: ::   |||   ::: | :: : ::   |  :|  ||
Qy      1 MSALGAAAPYLHHPADSHSGRVSFLGS-QPSPEVTAVAQLLKDLDRSTFRKLLKLVVGAL 59

Db    194 NAEDVRPVADRVTVQSAQIIDYQIRATLFVYPGPENEPIR-AAAEAKLKAYISAQHRLGR 252
          :: | |   :::   || : : ::   |        :::| ::|  |  |:    : ||
Qy     60 HGKDCREAVEQLGA-SANLSEERLAVLLAGTHTLLQQALRLPPASLKPDAFQEELQELGI 118

Db    253 DIRLSAIYAALHVEGVQRVELAAPVAD 279
          | :  |:|   | ||  |  :||:
Qy    119 PQDLIGDLASLAF-GSQR-PLLDSVAQ 143


RESULT   14
ID   Q9X7Y2      PRELIMINARY;      PRT;   382 AA.
AC   Q9X7Y2;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   PROBABLE ACYL-COA DEHYDROGENASE.
GN   SC6A5.36.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   OLIVER K., HARRIS D.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   BENTLEY S.D., PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE; 97000351.
RA   REDENBACH M., KIESER H.M., DENAPAITE D., EICHNER A., CULLUM J.,
RA   KINASHI H., HOPWOOD D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
```

```
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
DR   EMBL; AL049485; CAB39720.1; -.
DR   PROSITE; PS00072; ACYL_COA_DH_1; 1.
DR   PROSITE; PS00073; ACYL_COA_DH_2; 1.
SQ   SEQUENCE   382 AA;  42143 MW;  4E2F94FE CRC32;

  Query Match           6.7%;  Score 104;  DB 2;  Length 382;
  Best Local Similarity  27.2%;  Pred. No. 1.35e+00;
  Matches   28;  Conservative   27;  Mismatches  41;  Indels   7;  Gaps   7;

Db    240 GNLAQERLAIAV-GAAAATEEILDITTRYVKEREAFGRPLAKLQHVRFEIAEMATEA-AV 297
          :||::||||: : |: :  :: | :   :|  :||   |  |   :  |:::|: | :
Qy     75 ANLSEERLAVLLAGTHTLLQQALRLPPASLK-PDAFQEELQELGIPQDLIGDLASLAFGS 133

Db    298 TRTFLDRCVTEHVAGRLDPVHASMAKWWA-TELQKRVVDRCLQ 339
           | :||  |::: :: | | | |  :|:    :      | ||
Qy    134 QRPLLDS-VAQQQGSSL-P-HVSYFRWRVDVAISTSAQSRSLQ 173


RESULT   15
ID   O70211      PRELIMINARY;      PRT;   603 AA.
AC   O70211;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN COMPLEX ACID-LABILE
DE   SUBUNIT.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FISHER;
RX   MEDLINE; 98121980.
RA   DELHANTY P.J., BAXTER R.C.;
RT   "Cloning and characterization of the rat gene for the acid-labile
RT   subunit of the insulin-like growth factor binding protein complex.";
RL   J. Mol. Endocrinol. 19:267-277(1997).
DR   EMBL; AF006203; AAC15252.1; -.
DR   HSSP; P23945; 1XUN.
DR   PFAM; PF00560; LRR; 19.
DR   PFAM; PF01463; LRRCT; 1.
DR   PFAM; PF01462; LRRNT; 1.
DR   PRINTS; PR00019; LEURICHRPT.
SQ   SEQUENCE   603 AA;  66924 MW;  74E63165 CRC32;

  Query Match           6.7%;  Score 104;  DB 11;  Length 603;
  Best Local Similarity  31.9%;  Pred. No. 1.35e+00;
  Matches   38;  Conservative   31;  Mismatches  36;  Indels  14;  Gaps  13;

Db    272 DLSHNRVAGLMEDTFPGLLGLHVLRLAHNAIASLRPRTFKDLHFLEELQLGHNRIRQLGE 331
          :|| :|:|  |:  |   ||  : |||:  : |||:| :| :   |:|| : :: | :|:
Qy     76 NLSEERLAVLLAGT-HTLLQ-QALRLP--P-ASLKPDAFQE-E-LQELGIPQDLIGDLAS 128

Db    332 RTFEGLGQLEVL-TLNDNQITEV-RVGAFSGLFNVAVMNLSGNCLRSLPERVFQGLDKL 388
          :| | :|  :| :: : | : : :|: |  . :||: : |:   |||   |:  | ||
Qy    129 LAF-G-SQRPLLDSVAQQQGSSLPHVSYFRWRVDVAI-STSAQS-RSLQPSVLMQL-KL 182


Search completed: Fri May 12 10:25:41 2000
Job time : 87 secs.
```

****

MPSRCH (TM)

****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 12 10:23:17 2000; MasPar time 34.78 Seconds
192.339 Million cell updates/sec
Tabular output not generated.

Title: >US-09-223-796-2
Description: (1-224) from US09223796.pep
Perfect Score: 1552
Sequence: 1 MSALGAAAPYLHHPADSHSG.....ALVLKEMAELEKKCERKLQD 224

Scoring table: PAM 150
Gap 11

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 47.784; Variance 95.999; scale 0.498

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	120	7.7	661	1	TRA1_HUMAN	TUMOR NECROSIS FACTOR	3.38e-03
2	114	7.3	302	1	VPJ_BPP2	BASEPLATE ASSEMBLY PRO	2.14e-02
3	112	7.2	677	1	BS4_MOUSE	BS4 PROTEIN.	3.92e-02
4	106	6.8	1170	1	SMC2_YEAST	CHROMOSOME SEGREGATION	2.30e-01
5	104	6.7	603	1	ALS_RAT	INSULIN-LIKE GROWTH FA	4.08e-01
6	104	6.7	886	1	YFIQ_ECOLI	HYPOTHETICAL 98.0 KD P	4.08e-01
7	103	6.6	262	1	NCAP_INSVN	NUCLEOCAPSID PROTEIN (	5.43e-01
8	102	6.6	556	1	HSCC_ECOLI	CHAPEONE PROTEIN HSCC	7.20e-01
9	103	6.6	566	1	YBAE_ECOLI	HYPOTHETICAL 65.0 KD P	5.43e-01
10	102	6.6	603	1	ALS_MOUSE	INSULIN-LIKE GROWTH FA	7.20e-01
11	102	6.6	605	1	ALS_PAPPA	INSULIN-LIKE GROWTH FA	7.20e-01
12	100	6.4	219	1	YM50_MYCTU	HYPOTHETICAL TRANSCRIP	1.26e+00
13	100	6.4	804	1	YBBP_ECOLI	HYPOTHETICAL 89.3 KD P	1.26e+00
14	99	6.4	842	1	PKL2_RAT	PROTEIN KINASE C-LIKE	1.66e+00
15	100	6.4	2156	1	RRPL_PUUMH	RNA-DIRECTED RNA POLYM	1.26e+00
16	98	6.3	605	1	ALS_HUMAN	INSULIN-LIKE GROWTH FA	2.18e+00
17	98	6.3	2670	1	YAQ5_SCHPO	PUTATIVE TRANSLATIONAL	2.18e+00
18	97	6.2	201	1	MGF_CHICK	MYELOMONOCYTIC GROWTH	2.87e+00
19	97	6.2	584	1	YMB3_YEAST	HYPOTHETICAL 66.8 KD P	2.87e+00
20	97	6.2	1201	1	MFD_MYXXA	TRANSCRIPTION-REPAIR C	2.87e+00
21	96	6.2	1230	1	SMC3_YEAST	CHROMOSOME SEGREGATION	3.75e+00
22	95	6.1	63	1	YRKD_BACSU	HYPOTHETICAL 7.0 KD PR	4.91e+00
23	94	6.1	92	1	VNST_BUNSH	NONSTRUCTURAL PROTEIN	6.40e+00
24	94	6.1	314	1	FMT_PSEAE	METHIONYL-TRNA FORMYLT	6.40e+00
25	94	6.1	606	1	PRIM_MYXXA	DNA PRIMASE (EC 2.7.7.	6.40e+00
26	94	6.1	676	1	TIM_DROHY	TIMELESS PROTEIN (FRAG	6.40e+00
27	94	6.1	691	1	SLP1_YEAST	SLP1 PROTEIN (VACUOLAR	6.40e+00
28	94	6.1	807	1	YAK1_YEAST	PROTEIN KINASE YAK1 (E	6.40e+00
29	95	6.1	912	1	UBP3_YEAST	UBIQUITIN CARBOXYL-TER	4.91e+00
30	94	6.1	1363	1	XDH_EMENI	XANTHINE DEHYDROGENASE	6.40e+00
31	95	6.1	4568	1	DYHB_CHLRE	DYNEIN BETA CHAIN, FLA	4.91e+00
32	93	6.0	269	1	KAF1_SORBI	KAFIRIN PGK1 PRECURSOR	8.32e+00
33	93	6.0	308	1	NOSF_PSEST	COPPER TRANSPORT ATP-B	8.32e+00
34	93	6.0	315	1	ACCA_HAEIN	ACETYL-COENZYME A CARB	8.32e+00
35	93	6.0	390	1	PROB_MEIRU	GLUTAMATE 5-KINASE (EC	8.32e+00
36	93	6.0	456	1	PUR8_HAEIN	ADENYLOSUCCINATE LYASE	8.32e+00
37	93	6.0	727	1	GPDM_MOUSE	GLYCEROL-3-PHOSPHATE D	8.32e+00
38	93	6.0	887	1	RB6K_MOUSE	RAB6-INTERACTING KINES	8.32e+00
39	93	6.0	1106	1	GLI1_HUMAN	ZINC FINGER PROTEIN GL	8.32e+00
40	92	5.9	384	1	YEAP_ECOLI	HYPOTHETICAL 43.6 KD P	1.08e+01
41	92	5.9	452	1	PIV2_ADE12	MATURATION PROTEIN (PR	1.08e+01
42	92	5.9	490	1	MODF_ECOLI	PUTATIVE MOLYBDENUM TR	1.08e+01
43	92	5.9	498	1	YIEN_ECOLI	HYPOTHETICAL 56.4 KD P	1.08e+01
44	92	5.9	1391	1	N155_HUMAN	NUCLEAR PORE COMPLEX P	1.08e+01
45	92	5.9	2444	1	NTC1_HUMAN	NEUROGENIC LOCUS NOTCH	1.08e+01

ALIGNMENTS

```
RESULT    1
ID   TRA1_HUMAN     STANDARD;      PRT;   661 AA.
AC   Q12931; O75235;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   TUMOR NECROSIS FACTOR TYPE 1 RECEPTOR ASSOCIATED PROTEIN (TRAP-1)
DE   (FRAGMENT).
GN   TRAP1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95181307.
RA   SONG H.Y., DUNBAR J.D., ZHANG Y.X., GUO D., DONNER D.B.;
RT   "Identification of a protein with homology to hsp90 that binds the
RT   type 1 tumor necrosis factor receptor.";
RL   J. Biol. Chem. 270:3574-3581(1995).
RN   [2]
RP   SEQUENCE OF 16-631 FROM N.A.
RA   RICKE D.O., BRUCE D., MUNDT M., DOGGETT N., MUNK C., SAUNDERS E.,
RA   ROBINSON D., JONES M., BUCKINGHAM J., CHASTEEN L., THOMPSON S.,
RA   GOODWIN L., BRYANT J., TESMER J., MEINCKE L., LONGMIRE J., WHITE S.,
RA   UENG S., TATUM O., CAMPBELL C., FAWCETT J., MALTBIE M., MISRA M.,
RA   DEAVEN L.;
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: BINDS TO THE INTRACELLULAR DOMAIN OF TUMOR NECROSIS
CC       FACTOR TYPE 1 RECEPTOR.
CC   -!- TISSUE SPECIFICITY: FOUND IN SKELETAL MUSCLE, LIVER, HEART, BRAIN,
CC       KINDEY, PANCREAS, LUNG AND PLACENTA.
CC   -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U12595; AAA87704.1; -.
DR   EMBL; AC005203; AAC24722.1; -.
DR   PROSITE; PS00298; HSP90; FALSE_NEG.
DR   PFAM; PF00183; HSP90; 2.
FT   NON_TER       1      1
FT   CONFLICT    292    292       G -> R (IN REF. 2).
```

```
SQ   SEQUENCE   661 AA;  75342 MW;  055E95AF CRC32;

  Query Match             7.7%;  Score 120;  DB 1;  Length 661;
  Best Local Similarity  22.3%;  Pred. No. 3.38e-03;
  Matches    23;  Conservative   40;  Mismatches  34;  Indels   6;  Gaps   6;

Db      80 KKLLDIVARSLYS-E-KEVFIRELISNASDALEKLRHKLVSDGQALPEMEIHLQTNAEKG 137
           :||| :|  :| : : :|  : :| ::|: : |:|   |::  ::| :  ::|   : |
Qy      49 RKLLKLVVGALHGKDCREA-VEQLGASANLSEERLAV-LLAGTHTLLQQALRLPPASLKP 106

Db     138 TITIQD-TGIGMTQEELVSNLGTIARSGSKAFLDALQNQAEAS 179
             ::   :|: |: |:::|:::| :: :::||::  |  :|
Qy     107 DAFQEELQELGIPQD-LIGDLASLAFGSQRPLLDSVAQQQGSS 148


RESULT    2
ID   VPJ_BPP2       STANDARD;      PRT;   302 AA.
AC   P51767;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-DEC-1999 (Rel. 39, Last annotation update)
DE   BASEPLATE ASSEMBLY PROTEIN J (GPJ).
GN   J.
OS   Bacteriophage P2.
OC   Viruses; dsDNA viruses, no RNA stage; Tailed phages; Myoviridae.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96036485.
RA   HAGGARD-LJUNGQUIST E., JACOBSEN E., RISHOVD S., SIX E.W., NILSSEN O.,
RA   SUNSHINE M.G., LINDQVIST B.H., KIM K.-J., BARREIRO V., KOONIN E.V.,
RA   CALENDAR R.;
RT   "Bacteriophage P2: genes involved in baseplate assembly.";
RL   Virology 213:109-121(1995).
CC   -!- FUNCTION: LIES AT THE EDGE OF THE BASEPLATE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF063097; AAD03284.1; -.
SQ   SEQUENCE   302 AA;  32780 MW;  CB8BBA88 CRC32;

  Query Match             7.3%;  Score 114;  DB 1;  Length 302;
  Best Local Similarity  25.9%;  Pred. No. 2.14e-02;
  Matches    37;  Conservative   35;  Mismatches  66;  Indels   5;  Gaps   5;

Db     135 LSVAGSVGAYQYH-GRSADGRVADISVTSPSPACVTISVLSRENNGVASEDLLAVVRNAL 193
           :|  |: ::|  | : |  |||: :: : |||   ::: | :: :  :   || :|  ||
Qy       1 MSALGAAAPYLHHPADSHSGRVSFLG-SQPSPEVTAVAQLLKDLDRSTFRKLLKLVVGAL 59

Db     194 NGEDVRPVADRVTVQSAAIVEYQINATLYLYPGPESEPIR-AAAVKKLEAYITAQHRLGR 252
           :| | |   :::   || : | ::   |        :::| ::|  | :|:     : ||
Qy      60 HGKDCREAVEQLGA-SANLSEERLAVLLAGTHTLLQQALRLPPASLKPDAFQEELQELGI 118

Db     253 DIRLSAIYAALHVEGVQRVELAA 275
             | :  |:|   | ||  | :
Qy     119 PQDLIGDLASLAF-GSQRPLLDS 140


RESULT    3
ID   BS4_MOUSE      STANDARD;      PRT;   677 AA.
AC   P54729;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   BS4 PROTEIN.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=CONNECTIVE TISSUE;
RA   SHAN B., PARSA A.T., LEWIS J.A.;
RL   Submitted (JUN-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- INDUCTION: BY INTERFERON.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U27462; AAA68612.1; -.
DR   PFAM; PF00627; UBA; 3.
SQ   SEQUENCE   677 AA;  77448 MW;  A0235186 CRC32;

  Query Match             7.2%;  Score 112;  DB 1;  Length 677;
  Best Local Similarity  28.5%;  Pred. No. 3.92e-02;
  Matches    39;  Conservative   29;  Mismatches  59;  Indels  10;  Gaps  10;

Db     401 GNVDHAATHISNRREELAQIRKE-EKEKRRRRLE-NVNTLRGMGYSTQAAKQALHQARGN 458
           | |   ::: |     :||: |: ::   |: |   | :|:|:    :|  | |  | :|
Qy      20 GRVSFLGSQPSPEVTAVAQLLKDLDRSTFRKLLKLVVGALHGKDCR-EAVEQ-L-GASAN 76

Db     459 L-DDALKVLLSNPHMWWLQDADP-ENNSRQASPSQESINQLVYMGFDTVVAEAALRVFGG 516
           | :: | |||:  |   ||:|      |  : : || : :|  :  | :   |:|  ||:
Qy      77 LSEERLAVLLAGTHTL-LQQALRLPPASLKPDAFQEELQELG-IPQDLIGDLASL-AFGS 133

Db     517 NVQLAAQTLAHHGGSLP 533
             |      ::|:|||
Qy     134 QRPLLDSVAQQQGSSLP 150


RESULT    4
ID   SMC2_YEAST     STANDARD;      PRT;  1170 AA.
AC   P38989;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   CHROMOSOME SEGREGATION PROTEIN SMC2 (DA-BOX PROTEIN SMC2).
GN   SMC2 OR YFR031C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95212908.
RA   STRUNNIKOV A.V., HOGAN E., KOSHLAND D.;
RT   "SMC2, a Saccharomyces cerevisiae gene essential for chromosome
RT   segregation and condensation, defines a subgroup within the SMC
RT   family.";
RL   Genes Dev. 9:587-599(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RX   MEDLINE; 95400292.
RA   MURAKAMI Y., NAITOU M., HAGIWARA H., SHIBATA T., OZAWA M.,
RA   SASANUMA S.-I., SASANUMA M., TSUCHIYA Y., SOEDA E., YOKOYAMA K.,
RA   YAMAZAKI M., TASHIRO H., EKI T.;
RT   "Analysis of the nucleotide sequence of chromosome VI from
RT   Saccharomyces cerevisiae.";
RL   Nat. Genet. 10:261-268(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RX   MEDLINE; 96287654.
RA   EKI T., NAITOU M., HAGIWARA H., ABE M., OZAWA M., SASANUMA S.-I.,
RA   SASANUMA M., TSUCHIYA Y., SHIBATA T., WTANABE K., ONO A.,
RA   YAMAZAKI M.-A., TASHIRO H., HANAOKA F., MURAKAMI Y.;
RT   "Fifteen open reading frames in a 30.8 kb region of the right arm of
```

```
RT   chromosome VI from Saccharomyces cerevisiae.";
RL   Yeast 12:177-190(1996).
CC   -!- FUNCTION: INVOLVED IN CHROMOSOME SEGREGATION IN MITOSIS. COULD BE
CC       PART OF A CHROMOSOME CONDENSATION MOTOR.
CC   -!- SUBUNIT: HOMODIMER OR HETERODIMER WITH SMC1 OR OLIGOMERS.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- DOMAIN: CONSISTS OF TWO PUTATIVE CENTRAL COILED-COIL REGIONS
CC       FLANKED BY PUTATIVE GLOBULAR REGIONS AT THE N- AND C-TERMINUS.
CC   -!- SIMILARITY: BELONGS TO THE SMC FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U05820; AAA17416.1; -.
DR   EMBL; D50617; BAA09270.1; -.
DR   EMBL; D44602; BAA08042.1; -.
DR   PIR; S48530; S48530.
DR   HSSP; P02633; 1BOD.
DR   SGD; L0001927; SMC2.
KW   Mitosis; ATP-binding; Coiled coil; Nuclear protein.
FT   NP_BIND      32     39       ATP (POTENTIAL).
FT   DOMAIN      172    469       COILED COIL (POTENTIAL).
FT   DOMAIN      678   1027       COILED COIL (POTENTIAL).
FT   DOMAIN     1093   1119       ALA/ASP-RICH (DA-BOX).
SQ   SEQUENCE   1170 AA;  133927 MW;  C8865AC0 CRC32;

  Query Match             6.8%;  Score 106;  DB 1;  Length 1170;
  Best Local Similarity  26.9%;  Pred. No. 2.30e-01;
  Matches    25;  Conservative   21;  Mismatches  44;  Indels   3;  Gaps   3;

Db    796 LKKELKLLAKELEEQESESERKYDLFQNLELETEQLSSELDSNKTLLHNHLK-SIESLKL 854
          ::| |||:   |   :   |   :|  : :|  |:|:  | :  |||:  |: :  |||
Qy     48 FRKLLKLVVGALHGKDC-REAVEQLGASANLSEERLAVLLAGTHTLLQQALRLPPASLKP 106

Db    855 ENSDLEGKIRGVEDDLVTVQTELN-EEKKRLMD 886
          :  : |    |: :||:   : |     : |:|
Qy    107 DAFQEELQELGIPQDLIGDLASLAFGSQRPLLD 139


RESULT    5
ID   ALS_RAT        STANDARD;      PRT;   603 AA.
AC   P35859;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN COMPLEX ACID LABILE CHAIN
DE   PRECURSOR (ALS).
GN   IGFALS OR ALS.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 93038676.
RA   DAI J., BAXTER R.C.;
RT   "Molecular cloning of the acid-labile subunit of the rat insulin-like
RT   growth factor binding protein complex.";
RL   Biochem. Biophys. Res. Commun. 188:304-309(1992).
RN   [2]
RP   SEQUENCE OF 24-44, AND CHARACTERIZATION.
RC   STRAIN=WISTAR; TISSUE=SERUM;
RX   MEDLINE; 94130835.
RA   BAXTER R.C., DAI J.;
RT   "Purification and characterization of the acid-labile subunit of rat
RT   serum insulin-like growth factor binding protein complex.";
RL   Endocrinology 134:848-852(1994).
CC   -!- FUNCTION: MAY HAVE AN IMPORTANT ROLE IN REGULATING THE ACCESS OF
CC       CIRCULATING IGFS TO THE TISSUES.
CC   -!- SUBUNIT: FORMS A TERNARY COMPLEX OF ABOUT 140 TO 150 KD WITH IGF-I
CC       OR IGF-II AND IGFBP-3.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- TISSUE SPECIFICITY: BRAIN, KIDNEY, LUNG, HEART, SPLEEN, MUSCLE
CC       AND LIVER.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 20.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; S46785; CAB29017.1; -.
DR   PIR; JC1282; JC1282.
DR   HSSP; P23945; 1XUN.
DR   PFAM; PF00560; LRR; 19.
DR   PFAM; PF01462; LRRNT; 1.
DR   PFAM; PF01463; LRRCT; 1.
KW   Glycoprotein; Leucine-repeat; Repeat; Signal.
FT   SIGNAL        1     23
FT   CHAIN        24    603       INSULIN-LIKE GROWTH FACTOR BINDING
FT                                 PROTEIN COMPLEX ACID LABILE CHAIN.
FT   DOMAIN        79    535       LEUCINE-RICH REPEATS.
FT   REPEAT        79     89       LRR 1.
FT   REPEAT        90    113       LRR 2.
FT   REPEAT       114    137       LRR 3.
FT   REPEAT       138    161       LRR 4.
FT   REPEAT       162    185       LRR 5.
FT   REPEAT       186    209       LRR 6.
FT   REPEAT       210    233       LRR 7.
FT   REPEAT       234    257       LRR 8.
FT   REPEAT       258    281       LRR 9.
FT   REPEAT       282    305       LRR 10.
FT   REPEAT       306    329       LRR 11.
FT   REPEAT       330    353       LRR 12.
FT   REPEAT       354    377       LRR 13.
FT   REPEAT       378    401       LRR 14.
FT   REPEAT       402    425       LRR 15.
FT   REPEAT       426    449       LRR 16.
FT   REPEAT       450    473       LRR 17.
FT   REPEAT       474    497       LRR 18.
FT   REPEAT       498    521       LRR 19.
FT   REPEAT       522    535       LRR 20.
FT   CARBOHYD      64     64       POTENTIAL.
FT   CARBOHYD      85     85       POTENTIAL.
FT   CARBOHYD      96     96       POTENTIAL.
FT   CARBOHYD     368    368       POTENTIAL.
FT   CARBOHYD     515    515       POTENTIAL.
FT   CARBOHYD     578    578       POTENTIAL.
FT   CARBOHYD     586    586       POTENTIAL.
SQ   SEQUENCE   603 AA;  66811 MW;  5BB22D53 CRC32;

  Query Match             6.7%;  Score 104;  DB 1;  Length 603;
  Best Local Similarity  31.9%;  Pred. No. 4.08e-01;
  Matches    38;  Conservative   31;  Mismatches  36;  Indels  14;  Gaps  13;

Db    272 DLSHNRVAGLMEDTFPGLLGLHVLRLAHNAIASLRPRTFKDLHFLEELQLGHNRIRQLGE 331
          :|| :|:| |:  |   ||  : |||:  : |||:| :| :   |:|| : :: | :|:
Qy     76 NLSEERLAVLLAGT-HTLLQ-QALRLP--P-ASLKPDAFQE-E-LQELGIPQDLIGDLAS 128

Db    332 RTFEGLGQLEVL-TLNDNQITEV-RVGAFSGLFNVAVMNLSGNCLRSLPERVFQGLDKL 388
          :| | :|  :| :: : | : : :|: |    :||: : |:   |||   |:  | ||
Qy    129 LAF-G-SQRPLLDSVAQQQGSSLPHVSYFRWRVDVAI-STSAQS-RSLQPSVLMQL-KL 182


RESULT    6
ID   YFIQ_ECOLI     STANDARD;      PRT;   886 AA.
AC   P76594; Q47320;
```

```
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HYPOTHETICAL 98.0 KD PROTEIN IN UNG-PSSA INTERGENIC REGION.
GN   YFIQ.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE; 97426617.
RA   BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA   RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA   GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA   MAU B., SHAO Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE OF 1-612 FROM N.A.
RC   STRAIN=K12;
RA   NASHIMOTO H., SAITO N.;
RL   Submitted (SEP-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: SOME, IN THE N-TERMINAL, TO SUCCINYL-COA LIGASES.
CC   -!- SIMILARITY: STRONG, TO M.JANNASCHII MJ0590.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000344; AAC75637.1; -.
DR   EMBL; D64044; BAA10925.1; -.
DR   ECOGENE; EG14224; YFIQ.
DR   PFAM; PF00583; Acetyltransf; 1.
KW   Hypothetical protein.
FT   CONFLICT    506    507       ST -> YA (IN REF. 2).
SQ   SEQUENCE   886 AA;  97987 MW;  5D400142 CRC32;

  Query Match             6.7%;  Score 104;  DB 1;  Length 886;
  Best Local Similarity  27.2%;  Pred. No. 4.08e-01;
  Matches    22;  Conservative   19;  Mismatches  39;  Indels   1;  Gaps   1;

Db    416 RLFSEAGLPTYRTPEGTITAFMHMVEYRRNQKQLRETPALPSN-LTSNTAEAHLLLQQAI 474
          :|: :   :|:|     : : :|  : |    ||  :: |: : |:   | :| |||||:
Qy     38 QLLKDLDRSTFRKLLKLVVGALHGKDCREAVEQLGASANLSEERLAVLLAGTHTLLQQAL 97

Db    475 AEGATSLDTHEVQPILQAYGM 495
            ::||     |  ||  |:
Qy     98 RLPPASLKPDAFQEELQELGI 118


RESULT    7
ID   NCAP_INSVN     STANDARD;      PRT;   262 AA.
AC   Q01808;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-FEB-1994 (Rel. 28, Last annotation update)
DE   NUCLEOCAPSID PROTEIN (NUCLEOPROTEIN).
GN   N.
OS   Impatiens necrotic spot virus (strain NL-07) (INSV).
OC   Viruses; ssRNA negative-strand viruses; Bunyaviridae; Tospovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92331780.
RA   DE HAAN P., DE AVILA A.C., KORMELINK R., WESTERBROEK A.,
RA   GIELEN J.J.L., PETERS D., GOLDBACH R.;
RT   "The nucleotide sequence of the S RNA of Impatiens necrotic spot
RT   virus, a novel tospovirus.";
RL   FEBS Lett. 306:27-32(1992).
CC   -!- FUNCTION: ENCAPSIDATES THE RNA.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X66972; CAA47383.1; -.
DR   PIR; S23159; S23159.
DR   PFAM; PF01533; Tospo_nucleocap; 1.
KW   Nucleocapsid; RNA-binding.
SQ   SEQUENCE   262 AA;  28718 MW;  1B526B0F CRC32;

  Query Match             6.6%;  Score 103;  DB 1;  Length 262;
  Best Local Similarity  36.2%;  Pred. No. 5.43e-01;
  Matches    17;  Conservative   12;  Mismatches  17;  Indels   1;  Gaps   1;

Db    137 ATTDLKSCVLGVLLGGSLPLIASVLNFEIAALVPAIYQDAKHVELGI 183
          |:::|    |:|||:|:  |:   | :  |:| |  :|:  : ||||
Qy     73 ASANLSEERLAVLLAGTHTLLQQALRLPPASLKPDAFQEELQ-ELGI 118


RESULT    8
ID   HSCC_ECOLI     STANDARD;      PRT;   556 AA.
AC   P77319;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-DEC-1999 (Rel. 39, Last annotation update)
DE   CHAPEONE PROTEIN HSCC (HSC62).
GN   HSCC.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE; 97426617.
RA   BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA   RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA   GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA   MAU B., SHAO Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   CHUNG E., ALLEN E., ARAUJO R., APARICIO A., DAVIS K., DUNCAN M.,
RA   FEDERSPIEL N., HYMAN R., KALMAN S., KOMP C., KURDI O., LEW H., LIN D.,
RA   NAMATH A., OEFNER P., ROBERTS D., SCHRAMM S., DAVIS R.W.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE; 97061202.
RA   OSHIMA T., AIBA H., BABA T., FUJITA K., HAYASHI K., HONJO A.,
RA   IKEMOTO K., INADA T., ITOH T., KAJIHARA M., KANAI K., KASHIMOTO K.,
RA   KIMURA S., KITAGAWA M., MAKINO K., MASUDA S., MIKI T., MIZOBUCHI K.,
RA   MORI H., MOTOMURA K., NAKAMURA Y., NASHIMOTO H., NISHIO Y., SAITO N.,
RA   SAMPEI G., SEKI Y., TAGAMI H., TAKEMOTO K., WADA C., YAMAMOTO Y.,
RA   YANO M., HORIUCHI T.;
RT   "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 12.7-28.0 min region on the linkage map.";
RL   DNA Res. 3:137-155(1996).
RN   [4]
RP   CHARACTERIZATION.
RX   MEDLINE; 98407901.
RA   YOSHIMUNE K., YOSHIMURA T., ESAKI N.;
RT   "Hsc62, a new DnaK homologue of Escherichia coli.";
RL   Biochem. Biophys. Res. Commun. 250:115-118(1998).
CC   -!- FUNCTION: PROBABLE CHAPERONE. HAS ATPASE ACITIVITY. NOT STIMULATED
CC       BY DNAJ.
```

```
CC   -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000169; AAC73751.1; -.
DR   EMBL; U82598; AAB40851.1; -.
DR   EMBL; D90705; BAA35302.1; -.
DR   EMBL; D90704; BAA35297.1; -.
DR   HSSP; P19120; 1ATS.
DR   ECOGENE; EG13653; HSCC.
DR   PROSITE; PS00297; HSP70_1; 1.
DR   PROSITE; PS00329; HSP70_2; 1.
DR   PROSITE; PS01036; HSP70_3; 1.
DR   PFAM; PF00012; HSP70; 1.
KW   Chaperone; ATP-binding.
SQ   SEQUENCE   556 AA;  61986 MW;  44794D92 CRC32;

  Query Match             6.6%;  Score 102;  DB 1;  Length 556;
  Best Local Similarity  23.1%;  Pred. No. 7.20e-01;
  Matches    24;  Conservative   33;  Mismatches  43;  Indels   4;  Gaps   4;

Db     269 EDLWLPLLNRLRV-PIEQALRDARLKPSQIDSLVLVGGASQMPLVQRIAVRLFGKLPYQS 327
           : |:||    |:  :::: |::  :    |  |  :: :|| ||:: :| :  : ||  |
Qy      95 QALRLPP-ASLKPDAFQEELQELGIPQDLIGDLASLAFGSQRPLLDSVAQQQGSSLPHVS 153

Db     328 YDPSTI-VALGAAIQAACRLRSEDIEEVILTDICPYSLGVEVNR 370
           |    : ||:::: |:   |::  : :: ||| :  : | : :
Qy     154 YFRWRVDVAISTSAQSRS-LQPSVLMQLKLTDGSAHRFEVPIAK 196


RESULT    9
ID   YBAE_ECOLI     STANDARD;      PRT;   566 AA.
AC   P46890; P77158;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HYPOTHETICAL 65.0 KD PROTEIN IN HUPB-COF INTERGENIC REGION.
GN   YBAE.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE; 97426617.
RA   BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA   RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA   GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA   MAU B., SHAO Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RA   ROBERTS D., ALLEN E., ARAUJO R., APARICIO A., CHUNG E., DAVIS K.,
RA   DUNCAN M., FEDERSPIEL N., HYMAN R., KALMAN S., KOMP C., KURDI O.,
RA   LEW H., LIN D., NAMATH A., OEFNER P., SCHRAMM S., DAVIS R.W.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / W3110;
RA   HATADA E., OHMORI H., QIAO Y., TSUJI M., FUKUDA R.;
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 1-358 FROM N.A.
RC   STRAIN=K12 / MC4100;
RA   PATZER S.I., HANTKE K.;
RL   Submitted (OCT-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: TO E.COLI YABN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000151; AAC73548.1; -.
DR   EMBL; U82664; AAB40201.1; -.
DR   EMBL; D82943; BAA11649.1; -.
DR   EMBL; Z54355; CAA91182.1; -.
DR   ECOGENE; EG13217; YBAE.
DR   PFAM; PF00496; SBP_bac_5; 2.
KW   Hypothetical protein.
SQ   SEQUENCE   566 AA;  64991 MW;  091FA6F1 CRC32;

  Query Match             6.6%;  Score 103;  DB 1;  Length 566;
  Best Local Similarity  35.4%;  Pred. No. 5.43e-01;
  Matches    23;  Conservative   16;  Mismatches  19;  Indels   7;  Gaps   7;

Db      67 LRFLVTPESLRNAMMEQAL-ETGKQQDVL-ELAQLAPGELRTLLQPFMGGQWQNDTPTLR 124
           || |  | ||:    :: | | |  ||:: :|| || |  | ||:: :: | |  ::  :
Qy      97 LR-L-PPASLKPDAFQEELQELGIPQDLIGDLASLAFGSQRPLLDS-VA-Q-QQGSSLPH 151

Db     125 IPYYR 129
           ::|:|
Qy     152 VSYFR 156


RESULT   10
ID   ALS_MOUSE      STANDARD;      PRT;   603 AA.
AC   P70389;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN COMPLEX ACID LABILE CHAIN
DE   PRECURSOR (ALS).
GN   IGFALS OR ALS OR ALBS.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/SV;
RX   MEDLINE; 96413591.
RA   BOISCLAIR Y.R., SETO D., HSIEH S., HURST K.R., OOI G.T.;
RT   "Organization and chromosomal localization of the gene encoding the
RT   mouse acid labile subunit of the insulin-like growth factor binding
RT   complex.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:10028-10033(1996).
CC   -!- FUNCTION: MAY HAVE AN IMPORTANT ROLE IN REGULATING THE ACCESS OF
CC       CIRCULATING IGFS TO THE TISSUES.
CC   -!- SUBUNIT: FORMS A TERNARY COMPLEX OF ABOUT 140 TO 150 KD WITH IGF-I
CC       OR IGF-II AND IGFBP-3 (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 20.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U66900; AAB17270.1; -.
DR   MGD; MGI:107973; IGFALS.
DR   PFAM; PF00560; LRR; 19.
```

```
DR   PFAM; PF01462; LRRNT; 1.
DR   PFAM; PF01463; LRRCT; 1.
KW   Glycoprotein; Leucine-repeat; Repeat; Signal.
FT   SIGNAL        1     23       BY SIMILARITY.
FT   CHAIN        24    603       INSULIN-LIKE GROWTH FACTOR BINDING
FT                                 PROTEIN COMPLEX ACID LABILE CHAIN.
FT   DOMAIN        79    535       LEUCINE-RICH REPEATS.
FT   REPEAT        79     89       LRR 1.
FT   REPEAT        90    113       LRR 2.
FT   REPEAT       114    137       LRR 3.
FT   REPEAT       138    161       LRR 4.
FT   REPEAT       162    185       LRR 5.
FT   REPEAT       186    209       LRR 6.
FT   REPEAT       210    233       LRR 7.
FT   REPEAT       234    257       LRR 8.
FT   REPEAT       258    281       LRR 9.
FT   REPEAT       282    305       LRR 10.
FT   REPEAT       306    329       LRR 11.
FT   REPEAT       330    353       LRR 12.
FT   REPEAT       354    377       LRR 13.
FT   REPEAT       378    401       LRR 14.
FT   REPEAT       402    425       LRR 15.
FT   REPEAT       426    449       LRR 16.
FT   REPEAT       450    473       LRR 17.
FT   REPEAT       474    497       LRR 18.
FT   REPEAT       498    521       LRR 19.
FT   REPEAT       522    535       LRR 20.
FT   CARBOHYD      64     64       POTENTIAL.
FT   CARBOHYD      85     85       POTENTIAL.
FT   CARBOHYD      96     96       POTENTIAL.
FT   CARBOHYD     368    368       POTENTIAL.
FT   CARBOHYD     515    515       POTENTIAL.
FT   CARBOHYD     578    578       POTENTIAL.
FT   CARBOHYD     586    586       POTENTIAL.
SQ   SEQUENCE   603 AA;  66959 MW;  11ADB606 CRC32;

  Query Match           6.6%;  Score 102;  DB 1;  Length 603;
  Best Local Similarity  31.1%;  Pred. No. 7.20e-01;
  Matches   37;  Conservative   28;  Mismatches  40;  Indels  14;  Gaps  13;

Db    272 DLSHNRVAGLLEDTFPGLLGLHVLRLAHNAITSLRPRTFKDLHFLEELQLGHNRIRQLGE 331
          :|| :|:| ||  |   ||  : |||:  |  ||:| :| :   |:|| : :: | :|:
Qy     76 NLSEERLAVLLAGT-HTLLQ-QALRLPP-A--SLKPDAFQE-E-LQELGIPQDLIGDLAS 128

Db    332 KTFEGLGQLEVL-TLNDNQIHEVKVGAFFGLFNVAV-MNLSGNCLRSLPEHVFQQGLGRL 388
           :| | :|  :| :: : |   :   ::|    | | :: |:   |||   |:  | :|
Qy    129 LAF-G-SQRPLLDSVAQQQGSSLPHVSYFR-WRVDVAISTSAQS-RSLQPSVLMQL-KL 182


RESULT   11
ID   ALS_PAPPA      STANDARD;      PRT;   605 AA.
AC   O02833;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN COMPLEX ACID LABILE CHAIN
DE   PRECURSOR (ALS).
GN   IGFALS OR ALS.
OS   Papio papio (Guinea baboon).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Cercopithecidae; Cercopithecinae;
OC   Papio.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 97040714.
RA   DELHANTY P., BAXTER R.C.;
RT   "The cloning and expression of the baboon acid-labile subunit of the
RT   insulin-like growth factor binding protein complex.";
RL   Biochem. Biophys. Res. Commun. 227:897-902(1996).
CC   -!- FUNCTION: INVOLVED IN PROTEIN-PROTEIN INTERACTIONS THAT RESULT
CC       IN PROTEIN COMPLEXES, RECEPTOR-LIGAND BINDING OR CELL ADHESION.
CC   -!- SUBUNIT: FORMS A TERNARY COMPLEX OF ABOUT 140 TO 150 KD WITH IGF-I
CC       OR IGF-II AND IGFBP-3 (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 20.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; S83462; CAB34103.1; -.
DR   HSSP; P23945; 1XUN.
DR   PFAM; PF00560; LRR; 19.
DR   PFAM; PF01462; LRRNT; 1.
DR   PFAM; PF01463; LRRCT; 1.
KW   Glycoprotein; Leucine-repeat; Repeat; Signal.
FT   SIGNAL        1     27       BY SIMILARITY.
FT   CHAIN        28    605       INSULIN-LIKE GROWTH FACTOR BINDING
FT                                 PROTEIN COMPLEX ACID LABILE CHAIN.
FT   DOMAIN        79    536       LEUCINE-RICH REPEATS.
FT   REPEAT        79     89       LRR 1.
FT   REPEAT        90    113       LRR 2.
FT   REPEAT       114    137       LRR 3.
FT   REPEAT       138    161       LRR 4.
FT   REPEAT       162    185       LRR 5.
FT   REPEAT       186    209       LRR 6.
FT   REPEAT       210    233       LRR 7.
FT   REPEAT       234    257       LRR 8.
FT   REPEAT       258    281       LRR 9.
FT   REPEAT       282    305       LRR 10.
FT   REPEAT       306    329       LRR 11.
FT   REPEAT       330    353       LRR 12.
FT   REPEAT       354    377       LRR 13.
FT   REPEAT       378    401       LRR 14.
FT   REPEAT       402    425       LRR 15.
FT   REPEAT       426    449       LRR 16.
FT   REPEAT       450    473       LRR 17.
FT   REPEAT       474    497       LRR 18.
FT   REPEAT       498    521       LRR 19.
FT   REPEAT       522    536       LRR 20.
FT   CARBOHYD      64     64       POTENTIAL.
FT   CARBOHYD      85     85       POTENTIAL.
FT   CARBOHYD      96     96       POTENTIAL.
FT   CARBOHYD     368    368       POTENTIAL.
FT   CARBOHYD     515    515       POTENTIAL.
FT   CARBOHYD     580    580       POTENTIAL.
SQ   SEQUENCE   605 AA;  66110 MW;  5DF04D42 CRC32;

  Query Match           6.6%;  Score 102;  DB 1;  Length 605;
  Best Local Similarity  32.8%;  Pred. No. 7.20e-01;
  Matches   39;  Conservative   29;  Mismatches  37;  Indels  14;  Gaps  13;

Db    272 DLSHNRVAGLLEDTFPGLLGLRVLRLSHNAIASLRPRTFEDLHFLEELQLGHNRIRQLAE 331
          :|| :|:| ||  |   ||  : |||:  : |||:| :|::   |:|| : :: | :||
Qy     76 NLSEERLAVLLAGT-HTLLQ-QALRLP--P-ASLKPDAFQE-E-LQELGIPQDLIGDLAS 128

Db    332 RSFEGLGQLEVL-TLDHNQLQEV-KVGAFLGLTNVAVMNLSGNCLRNLPEQVFRGLGKL 388
           :| | :|  :| :: : |   :  |: |    :||: : |:   |:|   |:  | ||
Qy    129 LAF-G-SQRPLLDSVAQQQGSSLPHVSYFRWRVDVAI-STSAQS-RSLQPSVLMQL-KL 182


RESULT   12
ID   YM50_MYCTU     STANDARD;      PRT;   219 AA.
AC   Q10528;
DT   01-OCT-1996 (Rel. 34, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-DEC-1999 (Rel. 39, Last annotation update)
DE   HYPOTHETICAL TRANSCRIPTIONAL REGULATOR RV2250C.
GN   RV2250C OR MTCY427.32C.
OS   Mycobacterium tuberculosis.
```

```
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE; 98295987.
RA   COLE S.T., BROSCH R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
RA   GORDON S.V., EIGLMEIER K., GAS S., BARRY C.E. III, TEKAIA F.,
RA   BADCOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,
RA   DAVIES R., DEVLIN K., FELTWELL T., GENTLES S., HAMLIN N., HOLROYD S.,
RA   HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L.,
RA   OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J.,
RA   RUTTER S., SEEGER K., SKELTON S., SQUARES S., SQARES R., SULSTON J.E.,
RA   TAYLOR K., WHITEHEAD S., BARRELL B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
CC   -!- SIMILARITY: BELONGS TO THE TETR/ACRR FAMILY OF TRANSCRIPTIONAL
CC       REGULATORS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z70692; CAA94669.1; -.
DR   PROSITE; PS01081; HTH_TETR_FAMILY; FALSE_NEG.
DR   PFAM; PF00440; tetR; 1.
KW   Hypothetical protein; Transcription regulation; DNA-binding.
FT   DNA_BIND      62     81       H-T-H MOTIF (BY SIMILARITY).
SQ   SEQUENCE   219 AA;  23875 MW;  F3F30AD6 CRC32;

  Query Match             6.4%;  Score 100;  DB 1;  Length 219;
  Best Local Similarity  25.6%;  Pred. No. 1.26e+00;
  Matches    23;  Conservative   20;  Mismatches  42;  Indels   5;  Gaps   5;

Db     54 VDYGVDRVTLAEIARRAGVSRPTV-YRRWPDTRSIMASMLTSHI-ADVLREVPLDGDDRE 111
          : :| :|  |  :|:: | | | | | ||    :| :|  :  : : || :: |
Qy    129 LAFGSQRPLLDSVAQQQGSSLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQLKLTDGSAH 188

Db    112 ALVKQIVAVADRLR-GDDLIMSVMHSELAR 140
          :   |    : || :  |::  | :|| :
Qy    189 RFEVPIAKFQE-LRYSVALVLKEM-AELEK 216


RESULT   13
ID   YBBP_ECOLI     STANDARD;      PRT;   804 AA.
AC   P77504;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HYPOTHETICAL 89.3 KD PROTEIN IN TESA-RHSD INTERGENIC REGION.
GN   YBBP.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE; 97426617.
RA   BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA   RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA   GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA   MAU B., SHAO Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   ROBERTS D., ALLEN E., ARAUJO R., APARICIO A., CHUNG E., DAVIS K.,
RA   DUNCAN M., FEDERSPIEL N., HYMAN R., KALMAN S., KOMP C., KURDI O.,
RA   LEW H., LIN D., NAMATH A., OEFNER P., SCHRAMM S., DAVIS R.W.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000155; AAC73598.1; -.
DR   EMBL; U82664; AAB40250.1; -.
DR   ECOGENE; EG13263; YBBP.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM     15     35       POTENTIAL.
FT   TRANSMEM    243    263       POTENTIAL.
FT   TRANSMEM    301    321       POTENTIAL.
FT   TRANSMEM    333    353       POTENTIAL.
FT   TRANSMEM    381    401       POTENTIAL.
FT   TRANSMEM    403    423       POTENTIAL.
FT   TRANSMEM    453    473       POTENTIAL.
FT   TRANSMEM    680    700       POTENTIAL.
FT   TRANSMEM    734    754       POTENTIAL.
FT   TRANSMEM    769    789       POTENTIAL.
SQ   SEQUENCE   804 AA;  89332 MW;  8CE9FA3F CRC32;

  Query Match             6.4%;  Score 100;  DB 1;  Length 804;
  Best Local Similarity  37.3%;  Pred. No. 1.26e+00;
  Matches    25;  Conservative   16;  Mismatches  22;  Indels   4;  Gaps   4;

Db    425 LLNVLRRMTLKSLPLRLAVSRLL-RQPWSTLSQLSAFS-LSFMLLALLLVLRGDLLDRWQ 482
          ||: | | |:: | |:| |: |  :: : :: ||:| : ||   ||:||     ||::
Qy     39 LLKDLDRSTFRKL-LKLVVGALHGKDCREAVEQLGASANLSEERLAVLLAGTHTLLQQ-A 96

Db    483 QQLPPES 489
            :||| |
Qy     97 LRLPPAS 103


RESULT   14
ID   PKL2_RAT       STANDARD;      PRT;   842 AA.
AC   O08874;
DT   15-DEC-1999 (Rel. 39, Created)
DT   15-DEC-1999 (Rel. 39, Last sequence update)
DT   15-DEC-1999 (Rel. 39, Last annotation update)
DE   PROTEIN KINASE C-LIKE 2 (EC 2.7.1.-) (PROTEIN-KINASE C-RELATED KINASE
DE   2) (PROTEASE-ACTIVATED KINASE 2) (PAK-2) (FRAGMENT).
GN   PRKCL2 OR PRK2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CHARACTERIZATION.
RC   TISSUE=MYELOMA, AND LIVER;
RX   MEDLINE; 97248559.
RA   YU W., LIU J., MORRICE N.A., WETTENHALL R.E.H.;
RT   "Isolation and characterization of a structural homologue of human
RT   PRK2 from rat liver: distinguishing substrate and lipid activator
RT   specificities.";
RL   J. Biol. Chem. 272:10030-10034(1997).
CC   -!- FUNCTION: EXHIBITS A PREFERENCE FOR HIGHLY BASIC PROTEIN
CC       SUBSTRATES.
CC   -!- ENZYME REGULATION: ACTIVATED BY LIPIDS, PARTICULARLY CARDIOLIPIN
CC       AND TO A LESSER EXTENT BY OTHER ACIDIC PHOSPHOLIPIDS AND
CC       UNSATURATED FATTY ACIDS.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- PTM: AUTOPHOSPHORYLATED.
CC   -!- PTM: ACTIVATED BY LIMITED PROTEOLYSIS WITH TRYPSIN.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC       PKC SUBFAMILY.
CC   ---------------------------------------------------------------------
```

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U75358; AAB53364.1; -.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PFAM; PF00069; pkinase; 1.
DR   PFAM; PF00433; pkinase_C; 1.
KW   Transferase; ATP-binding; Serine/threonine-protein kinase;
KW   Phosphorylation.
FT   NON_TER       1      1
FT   DOMAIN      521    780       PROTEIN KINASE.
FT   NP_BIND     527    535       ATP (BY SIMILARITY).
FT   BINDING     550    550       ATP (BY SIMILARITY).
FT   ACT_SITE    646    646       BY SIMILARITY.
FT   NON_TER     842    842
SQ   SEQUENCE   842 AA;  95650 MW;  B4B0C8E1 CRC32;

  Query Match             6.4%;  Score 99;  DB 1;  Length 842;
  Best Local Similarity  25.8%;  Pred. No. 1.66e+00;
  Matches    24;  Conservative   30;  Mismatches  34;  Indels   5;  Gaps   4;

Db      59 LQAVLTNELAFDNAKPVISPLELRNGRIIEH--HFRIEFAVAEGAKNVMKLLGSGKVTDR 116
           | : |:: ||| : :|:: ::  : |  : |   ||    || ::    : | :: : :
Qy     122 LIGDLAS-LAFGSQRPLLDSVAQQQGSSLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQL 180

Db     117 KALSEAQA-RFNESSQKLDLLKYSLEQRLNELP 148
           | |::: | ||: :  |:: |:||:   |:|::
Qy     181 K-LTDGSAHRFEVPIAKFQELRYSVALVLKEMA 212


RESULT   15
ID   RRPL_PUUMH     STANDARD;      PRT;  2156 AA.
AC   P27176;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   01-AUG-1992 (Rel. 23, Last annotation update)
DE   RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (L PROTEIN).
GN   L.
OS   Puumala virus (strain Hallnas B1) (Nephropathia epidemica virus).
OC   Viruses; ssRNA negative-strand viruses; Bunyaviridae; Hantavirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91272501.
RA   STOHWASSER R., RAAB K., DARAI G., BAUTZ E.K.F.;
RT   "Primary structure of the large (L) RNA segment of nephropathia
RT   epidemica virus strain Hallnas B1 coding for the viral RNA
RT   polymerase.";
RL   Virology 183:386-391(1991).
CC   -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE
CC       + RNA(N).
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M63194; -; NOT_ANNOTATED_CDS.
DR   PIR; A40319; RRVUNE.
KW   Transferase; Nucleotidyltransferase; RNA-directed RNA polymerase.
SQ   SEQUENCE   2156 AA;  247048 MW;  DEFC118F CRC32;

  Query Match             6.4%;  Score 100;  DB 1;  Length 2156;
  Best Local Similarity  23.5%;  Pred. No. 1.26e+00;

  Matches    28;  Conservative   39;  Mismatches  43;  Indels   9;  Gaps   9;

Db    1855 LLAHAYHNLKDVRYQALQAVGNVWFKMTQHKLFINPIISAGLLENFMKGLPAAIPPAAYS 1914
           |:  | |  || | :|:: :|    :::: :| :    :  ||:  ::  ||:: | |:
Qy      54 LVVGALHG-KDCR-EAVEQLGAS-ANLSEERLAVLLAGTHTLLQQALRLPPASLKPDAFQ 110

Db    1915 LIMNKAKISVDLFMFNELLAL-INPQNVLNLDGIEETSEGYTTVSTISSTQWSEEVSLT 1972
             :    |: || :  :| :| : :|  | ||:: :  :| :::: :|  :|  :|:::
Qy     111 EELQELGIPQDL-I-GDLASLAFGSQRPL-LDSVAQQ-QG-SSLPHVSYFRWRVDVAIS 164


Search completed: Fri May 12 10:23:56 2000
Job time : 39 secs.
```

****

MPSRCH (TM)

****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 12 10:25:59 2000; MasPar time 53.66 Seconds
196.913 Million cell updates/sec

Tabular output not generated.

Title: >US-09-223-796-2
Description: (1-224) from US09223796.pep
Perfect Score: 1552
Sequence: 1 MSALGAAAPYLHHPADSHSG.....ALVLKEMAELEKKCERKLQD 224

Scoring table: PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir62
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 46.684; Variance 106.313; scale 0.439

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	121	7.8	491	2	S75449	hypothetical protein	1.52e-02
2	120	7.7	661	2	A55877	tumor necrosis factor	2.01e-02
3	118	7.6	676	2	S41022	hypothetical protein	3.51e-02
4	110	7.1	706	2	S74610	hypothetical protein	3.12e-01
5	107	6.9	668	2	T11778	phosphoglycerate tran	6.92e-01
6	107	6.9	913	2	T15474	hypothetical protein	6.92e-01
7	106	6.8	1170	2	A56157	chromosome segregatio	8.99e-01
8	104	6.7	228	2	E72540	hypothetical protein	1.51e+00
9	104	6.7	603	2	JC1282	insulin-like growth f	1.51e+00
10	104	6.7	886	2	G65036	hypothetical protein	1.51e+00
11	104	6.7	1116	2	S77213	DNA-directed DNA poly	1.51e+00
12	102	6.6	159	2	T05656	hypothetical protein	2.52e+00
13	103	6.6	262	2	S23159	nucleoprotein N - Imp	1.96e+00
14	102	6.6	451	2	S74728	hypothetical protein	2.52e+00
15	103	6.6	555	2	F72111	fructose-6-p phosphot	1.96e+00
16	102	6.6	556	2	H64799	probable dnaK-type mo	2.52e+00
17	103	6.6	566	2	E64774	probable membrane pro	1.96e+00
18	102	6.6	603	2	JC6128	insulin-like growth f	2.52e+00
19	102	6.6	605	2	JC5239	insulin-like growth f	2.52e+00
20	100	6.4	219	2	F70779	hypothetical protein	4.19e+00
21	100	6.4	362	2	D72734	probable histidine de	4.19e+00
22	99	6.4	371	2	S76863	hypothetical protein	5.38e+00
23	100	6.4	804	2	G64780	probable membrane pro	4.19e+00
24	100	6.4	2156	1	RRVUNE	genome polyprotein -	4.19e+00
25	98	6.3	229	2	F70962	probable transcriptio	6.89e+00
26	98	6.3	336	2	S13493	alpha-1-microglobulin	6.89e+00
27	98	6.3	344	2	E70851	probable parA protein	6.89e+00
28	98	6.3	605	2	A41915	insulin-like growth f	6.89e+00
29	97	6.2	201	2	A42247	myelomonocytic growth	8.82e+00
30	96	6.2	312	2	S77670	probable transcriptio	1.13e+01
31	96	6.2	482	1	A34671	triacylglycerol lipas	1.13e+01
32	97	6.2	584	2	S55106	probable membrane pro	8.82e+00
33	97	6.2	669	2	S75785	translation elongatio	8.82e+00
34	96	6.2	1230	2	S56850	SMC1 protein homolog	1.13e+01
35	95	6.1	63	2	B69976	hypothetical protein	1.44e+01
36	95	6.1	287	2	D70860	hypothetical protein	1.44e+01
37	95	6.1	315	2	F71377	conserved hypothetica	1.44e+01
38	95	6.1	330	2	T14806	hypothetical protein	1.44e+01
39	95	6.1	452	2	T05406	hypothetical protein	1.44e+01
40	95	6.1	603	2	T16655	hypothetical protein	1.44e+01
41	94	6.1	606	2	S70829	primase - Myxococcus	1.83e+01
42	94	6.1	807	2	A32582	protein kinase YAK1 (	1.83e+01
43	95	6.1	912	2	B44450	ubiquitin-specific pr	1.44e+01
44	94	6.1	1363	1	A55875	xanthine dehydrogenas	1.83e+01
45	95	6.1	4568	2	T08030	dynein beta heavy cha	1.44e+01

ALIGNMENTS

RESULT 1
ENTRY S75449 #type complete
TITLE hypothetical protein sll1424 - Synechocystis sp. (strain PCC 6803)
ORGANISM #formal_name Synechocystis sp.
#variety PCC 6803
DATE 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 21-Aug-1998
ACCESSIONS S75449
REFERENCE S74322
#authors Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirosawa, M.; Sugiura, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.
#journal DNA Res. (1996) 3:109-136
#title Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.
#cross-references MUID:97061201
#accession S75449
##status nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-491 ##label KAN
##cross-references EMBL:D90911; GB:AB001339; NID:g1653083; PID:d1018743; PID:g1653094
##note the nucleotide sequence was submitted to the EMBL Data Library, June 1996
SUMMARY #length 491 #molecular-weight 55645 #checksum 8967

Query Match 7.8%; Score 121; DB 2; Length 491;
Best Local Similarity 23.1%; Pred. No. 1.52e-02;
Matches 27; Conservative 36; Mismatches 47; Indels 7; Gaps 7;

```
Db    121 RELNRN-YTKALQLLRRVSEQSRNLRQEVANLSAERAQLNEQK-DSLLAESSELQSQVKL 178
          ::|:|: : | | |:   : :::: |: |  |:|  | |:|::   ||| :  |  |  |
Qy     41 KDLDRSTFRKLLKLVV-GALHGKDCREAVEQLGAS-ANLSEERLAVLLAGTHTLLQQA-L 97

Db    179 RDQELSKRQERIAQQEKVLARQREQVQSLEKRFASLEAQRQQLQAEINQRDTKIDQL 235
          |    | : : : ::   |:  :: :  |   :| : :||  |::   |: : : ::
Qy     98 RLPPASLKPDAFQEELQELGIPQDLIGDLAS-LA-FGSQRPLLDSVAQQQGSSLPHV 152
```

RESULT 2
ENTRY A55877 #type complete

```
TITLE           tumor necrosis factor type 1 receptor associated protein
                  TRAP-1 - human
ORGANISM        #formal_name Homo sapiens #common_name man
DATE            23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change
                  26-Aug-1999
ACCESSIONS      A55877
REFERENCE       A55877
   #authors     Song, H.Y.; Dunbar, J.D.; Zhang, Y.X.; Guo, D.; Donner, D.B.
   #journal     J. Biol. Chem. (1995) 270:3574-3581
   #title       Identification of a protein with homology to hsp90 that binds
                  the type 1 tumor necrosis factor receptor.
   #cross-references MUID:95181307
   #accession   A55877
      ##status       preliminary
      ##molecule_type mRNA
      ##residues     1-661 ##label SON
      ##cross-references GB:U12595; NID:g687236; PID:g687237
CLASSIFICATION  #superfamily heat shock protein 90
SUMMARY         #length 661  #molecular-weight 75342  #checksum 9066

  Query Match            7.7%;  Score 120;  DB 2;  Length 661;
  Best Local Similarity  22.3%;  Pred. No. 2.01e-02;
  Matches    23;  Conservative   40;  Mismatches  34;  Indels   6;  Gaps   6;

Db      80 KKLLDIVARSLYS-E-KEVFIRELISNASDALEKLRHKLVSDGQALPEMEIHLQTNAEKG 137
           :||| :|  :| : : :|  : :| ::|: : |:|   |::  ::| :  ::|   : |
Qy      49 RKLLKLVVGALHGKDCREA-VEQLGASANLSEERLAV-LLAGTHTLLQQALRLPPASLKP 106

Db     138 TITIQD-TGIGMTQEELVSNLGTIARSGSKAFLDALQNQAEAS 179
             ::   :|: |: |:::|:::| :: :::||::  |  :|
Qy     107 DAFQEELQELGIPQD-LIGDLASLAFGSQRPLLDSVAQQQGSS 148


RESULT    3
ENTRY           S41022     #type complete
TITLE           hypothetical protein T07C4.9 - Caenorhabditis elegans
ORGANISM        #formal_name Caenorhabditis elegans
DATE            06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change
                  09-Sep-1997
ACCESSIONS      S41022
REFERENCE       S41014
   #authors     Berks, M.
   #submission  submitted to the EMBL Data Library, January 1994
   #accession   S41022
      ##status       preliminary
      ##molecule_type DNA
      ##residues     1-676 ##label BER
      ##cross-references EMBL:Z29443; NID:g1067051; PID:g443836
GENETICS
   #introns     69/3; 161/1; 208/3; 227/1; 357/1; 505/3
CLASSIFICATION  #superfamily annexin repeat homology
FEATURE
   373-444           #domain annexin repeat homology #label AX1\
   445-516           #domain annexin repeat homology #label AX6\
   528-600           #domain annexin repeat homology #label AX3\
   604-675           #domain annexin repeat homology #label AX4
SUMMARY         #length 676  #molecular-weight 74968  #checksum 2702

  Query Match            7.6%;  Score 118;  DB 2;  Length 676;
  Best Local Similarity  25.7%;  Pred. No. 3.51e-02;
  Matches    37;  Conservative   31;  Mismatches  67;  Indels   9;  Gaps   8;

Db      59 C-ELVKESRSHEMLSEEQLATLYDCATTSTRLPNDFADDQEDLKNIIRYGELFKACHAIN 117
           | | | :   :   ||||:|| |   | | | |   :      ||    : | :    :|
Qy      64 CREAVEQLGASANLSEERLAVLL--AGTHTLLQQALRLPPASLKPDA-FQEELQE-LGI- 118

Db     118 STDFIQKSEDLK-DEEKVALERIVEQKLAESAKNEKDIAWNVNIVVANSYVAKSLRPVIN 176
           : |:|     |    ::  |: : :|  : | :   : | |:: :: |  ::||:| :
Qy     119 PQDLIGDLASLAFGSQRPLLDSVAQQQ-GSSLPHVSYFRWRVDVAISTSAQSRSLQPSVL 177

Db     177 IQMPTVGGDTN-FEFDIDSFAQFR 199
           :|:    | :: ||  |  | ::|
Qy     178 MQLKLTDGSAHRFEVPIAKFQELR 201


RESULT    4
ENTRY           S74610     #type complete
TITLE           hypothetical protein slr1103 - Synechocystis sp. (strain PCC
                  6803)
ORGANISM        #formal_name Synechocystis sp.
   #variety     PCC 6803
DATE            25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
                  29-Sep-1999
ACCESSIONS      S74610
REFERENCE       S74322
   #authors     Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
                  Nakamura, Y.; Miyajima, N.; Hirosawa, M.; Sugiura, M.;
                  Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
                  Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo,
                  S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
                  Yasuda, M.; Tabata, S.
   #journal     DNA Res. (1996) 3:109-136
   #title       Sequence analysis of the genome of the unicellular
                  cyanobacterium Synechocystis sp. PCC6803. II. Sequence
                  determination of the entire genome and assignment of
                  potential protein-coding regions.
   #cross-references MUID:97061201
   #accession   S74610
      ##status       nucleic acid sequence not shown; translation not shown
      ##molecule_type DNA
      ##residues     1-706 ##label KAN
      ##cross-references EMBL:D90900; GB:AB001339; NID:g1651768;
                        PIDN:BAA16762.1; PID:d1017495; PID:g1651835
      ##note         the nucleotide sequence was submitted to the EMBL Data
                      Library, June 1996
CLASSIFICATION  #superfamily hypothetical protein sll1895
SUMMARY         #length 706  #molecular-weight 79830  #checksum 5483

  Query Match            7.1%;  Score 110;  DB 2;  Length 706;
  Best Local Similarity  27.7%;  Pred. No. 3.12e-01;
  Matches    23;  Conservative   29;  Mismatches  26;  Indels   5;  Gaps   5;

Db     585 RRAIEQLATIGVNLSLDDFGIGL-GTLSCLQQ-FKIPAVKIH-ESFIKDLEQSPVNEAII 641
           | |:|||:: : ||| : ::: | || : ||| :::|:  :  ::|  :|::  : : :|
Qy      65 REAVEQLGA-SANLSEERLAVLLAGTHTLLQQALRLPPASLKPDAFQEELQELGIPQDLI 123

Db     642 TSIMTLGRKLGVRIISEGVETQQ 664
             : :|:   : | : ::|  ||
Qy     124 GDLASLAFG-SQRPLLDSVAQQQ 145


RESULT    5
ENTRY           T11778     #type complete
TITLE           phosphoglycerate transport regulatory protein pgtB -
                  Salmonella typhimurium
ORGANISM        #formal_name Salmonella typhimurium
DATE            16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change
                  16-Jul-1999
ACCESSIONS      T11778
REFERENCE       Z17339
   #authors     Yang, Y.L.; Goldrick, D.; Hong, J.S.
   #journal     J. Bacteriol. (1988) 170:4299-4303
   #title       Identification of the products and nucleotide sequences of
                  two regulatory genes involved in the exogenous induction of
                  phosphoglycerate transport in Salmonella typhimurium.
   #cross-references MUID:88314933
   #accession   T11778
      ##status       preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues     1-668 ##label YAN
      ##cross-references EMBL:M21279; NID:g858752; PID:g858753
      ##experimental_source strain LT2
SUMMARY         #length 668  #molecular-weight 75231  #checksum 5379

  Query Match            6.9%;  Score 107;  DB 2;  Length 668;
  Best Local Similarity  31.5%;  Pred. No. 6.92e-01;
```

```
  Matches    28;  Conservative   18;  Mismatches  37;  Indels   6;  Gaps   5;

Db      99 DKIERLSQGLQLAERRQLAVILQDSRTLLAELDNALYNMFLVREKVSELSARLTGSSDDF 158
           : :|:|: :  |:| | |||:|  ::||| :     |   :    |    | | ::|:
Qy      66 EAVEQLGASANLSEER-LAVLLAGTHTLLQQ-ALRLPPASLKPDAFQEELQEL-GIPQDL 122

Db     159 TTELNSLVQDFTWQQGTLLDQIEANQGDA 187
             :| ||   |  |:  ||| :   || :
Qy     123 IGDLASLA--FGSQR-PLLDSVAQQQGSS 148


RESULT    6
ENTRY             T15474     #type complete
TITLE             hypothetical protein C09B9.7 - Caenorhabditis elegans
ORGANISM          #formal_name Caenorhabditis elegans
DATE              20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change
                    20-Sep-1999
ACCESSIONS        T15474
REFERENCE         Z18355
   #authors       Bradshaw, H.
   #submission    submitted to the EMBL Data Library, February 1996
   #description   The sequence of C. elegans cosmid C09B9.
   #accession     T15474
      ##status         preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues       1-913 ##label BRA
      ##cross-references EMBL:U50069; NID:g1702977; PID:g1702978;
                          PIDN:AAB37561.1; GSPDB:GN00022; CESP:C09B9.7
      ##experimental_source strain Bristol N2; clone C09B9
GENETICS
   #gene          CESP:C09B9.7
   #map_position 4
   #introns       758/2; 864/2; 895/2
SUMMARY           #length 913  #molecular-weight 103125  #checksum 7779

  Query Match             6.9%;  Score 107;  DB 2;  Length 913;
  Best Local Similarity  23.2%;  Pred. No. 6.92e-01;
  Matches    26;  Conservative   36;  Mismatches  45;  Indels   5;  Gaps   5;

Db     211 IGLTTDEAMSDVNQILELVNSVESDQFKKHSIGFFRGILSSKHLERESRDPWLGKVLNVS 270
           :::   :: ::|  : :|:: ::   |:|  : :  | | :|   ||: :  ||   |:|
Qy      22 VSFLGSQPSPEVTAVAQLLKDLDRSTFRKL-LKLVVGALHGKDC-REAVEQ-LGASANLS 78

Db     271 EEHLASITSGLQPLIKASDHLETVNDKLMSVRPNALSRSLGYFQSFVTTMSS 322
           ||:|| : :| : |:  : :|   : |  : : : | : ||  | ::  ::|
Qy      79 EERLAVLLAGTHTLLQQALRLPPASLKPDAFQ-EEL-QELGIPQDLIGDLAS 128


RESULT    7
ENTRY             A56157     #type complete
TITLE             chromosome segregation protein SMC2 - yeast (Saccharomyces
                    cerevisiae)
ALTERNATE_NAMES   DA-box protein; protein R003; protein YFR031c; structural
                    maintenance of chromosomes 2
ORGANISM          #formal_name Saccharomyces cerevisiae
DATE              03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change
                    24-Sep-1999
ACCESSIONS        A56157; S56286; S62242; S63836; S48530
REFERENCE         A56157
   #authors       Strunnikov, A.V.; Hogan, E.; Koshland, D.
   #journal       Genes Dev. (1995) 9:587-599
   #title         SMC2, a Saccharomyces cerevisiae gene essential for
                    chromosome segregation and condensation, defines a subgroup
                    within the SMC family.
   #cross-references MUID:95212908
   #accession     A56157
      ##status         preliminary
      ##molecule_type DNA
      ##residues       1-1170 ##label STR
      ##cross-references GB:U05820; NID:g468039; PIDN:AAA17416.1; PID:g468040
REFERENCE         S56186
   #authors       Murakami, Y.; Naitou, M.; Hagiwara, H.; Shibata, T.; Ozawa,
                    M.; Sasanuma, S.I.; Sasanuma, M.; Tsuchiya, Y.; Soeda, E.;
                    Yokoyama, K.; Yamazaki, M.; Tashiro, H.; Eki, T.
   #submission    submitted to the EMBL Data Library, May 1995
   #description   Analysis of the nucleotide sequence of chromosome VI from
                    Saccaromyces cerevisiae.
   #accession     S56286
      ##molecule_type DNA
      ##residues       1-1170 ##label MUR
      ##cross-references EMBL:D50617; NID:g836685; PIDN:BAA09270.1;
                          PID:d1009911; PID:g836786; MIPS:YFR031c
REFERENCE         S62230
   #authors       Murakami, Y.
   #submission    submitted to the EMBL Data Library, December 1994
   #accession     S62242
      ##molecule_type DNA
      ##residues       1-1170 ##label MUW
      ##cross-references EMBL:D44602; NID:g893419; PIDN:BAA08042.1;
                          PID:d1008633; PID:g893426
REFERENCE         S63830
   #authors       Eki, T.; Naitou, M.; Hagiwara, H.; Abe, M.; Ozawa, M.;
                    Sasanuma, S.I.; Sasanuma, M.; Tsuchiya, Y.; Shibata, T.;
                    Watanabe, K.; Ono, A.; Yamazaki, M.A.; Tashiro, H.;
                    Hanaoka, F.; Murakami, Y.
   #journal       Yeast (1996) 12:177-190
   #title         Fifteen open reading frames in a 30.8 kb region of the right
                    arm of chromosome VI from Saccharomyces cerevisiae.
   #cross-references MUID:96287654
   #accession     S63836
      ##status         nucleic acid sequence not shown; translation not shown
      ##molecule_type DNA
      ##residues       1-1170 ##label EKI
      ##cross-references EMBL:D44602; NID:g893419; PIDN:BAA08042.1;
                          PID:d1008633; PID:g893426
      ##note           the nucleotide sequence was submitted to the EMBL Data
                        Library, July 1995
GENETICS
   #gene          SGD:SMC2
      ##cross-references SGD:S0001927; MIPS:YFR031c
   #map_position 6R
CLASSIFICATION    #superfamily conserved hypothetical P115 protein
KEYWORDS          transmembrane protein
FEATURE
   41-57               #domain transmembrane #status predicted #label TM1\
   1090-1106           #domain transmembrane #status predicted #label TM2
SUMMARY           #length 1170  #molecular-weight 133927  #checksum 2840

  Query Match             6.8%;  Score 106;  DB 2;  Length 1170;
  Best Local Similarity  26.9%;  Pred. No. 8.99e-01;
  Matches    25;  Conservative   21;  Mismatches  44;  Indels   3;  Gaps   3;

Db     796 LKKELKLLAKELEEQESESERKYDLFQNLELETEQLSSELDSNKTLLHNHLK-SIESLKL 854
           ::| |||:   |   :   |   :|  : :|  |:|:  | :  |||:  |: :  |||
Qy      48 FRKLLKLVVGALHGKDC-REAVEQLGASANLSEERLAVLLAGTHTLLQQALRLPPASLKP 106

Db     855 ENSDLEGKIRGVEDDLVTVQTELN-EEKKRLMD 886
           :  : |    |: :||:   : |     : |:|
Qy     107 DAFQEELQELGIPQDLIGDLASLAFGSQRPLLD 139


RESULT    8
ENTRY             E72540     #type complete
TITLE             hypothetical protein APE1610 - Aeropyrum pernix (strain K1)
ORGANISM          #formal_name Aeropyrum pernix
DATE              20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change
                    20-Aug-1999
ACCESSIONS        E72540
REFERENCE         A72450
   #authors       Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.;
                    Haikawa, Y.; Jin-no, K.; Takahashi, M.; Sekine, M.; Baba,
                    S.; Ankai, A.; Kosugi, H.; Hosoyama, A.; Fukui, S.; Nagai,
                    Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
                    Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.;
                    Kushida, N.; Oguchi, A.; Aoki, K.; Kubota, K.; Nakamura,
                    Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
```

```
   #journal      DNA Res. (1999) 6:83-101
   #title        Complete genome sequence of an aerobic hyper-thermophilic
                   Crenarchaeon, Aeropyrum pernix K1.
   #cross-references MUID:99310339
   #accession    E72540
      ##status       preliminary
      ##molecule_type DNA
      ##residues      1-228 ##label KAW
      ##cross-references DDBJ:AP000062; NID:g5105244; PIDN:BAA80610.1;
                        PID:d1044396; PID:g5105297
      ##experimental_source strain K1
GENETICS
   #gene         APE1610
SUMMARY          #length 228  #molecular-weight 25890  #checksum 7192

  Query Match             6.7%;  Score 104;  DB 2;  Length 228;
  Best Local Similarity  32.2%;  Pred. No. 1.51e+00;
  Matches    29;  Conservative   22;  Mismatches  33;  Indels   6;  Gaps   6;

Db    131 VVQLVEKDVKQSPLRRKLQRLKLPDLAAEAYATEGDALAVARELWG-RLARLGAVFRVSI 189
          | ||: ||: :| :| || :| :  | :       : |: : :|   ||| | |  : ::
Qy     36 VAQLL-KDLDRSTFR-KLLKLVVGALHGKDCREAVEQLGASANLSEERLAVLLAGTH-TL 92

Db    190 VPSRRLVLAVASPRPDSFGPVVEEVLSIVQ 219
          :  : | |: || :||:|   ::| |:| |
Qy     93 LQ-QALRLPPASLKPDAFQEELQE-LGIPQ 120


RESULT    9
ENTRY            JC1282     #type complete
TITLE            insulin-like growth factor-binding protein acid labile chain
                   precursor - rat
ORGANISM         #formal_name Rattus norvegicus #common_name Norway rat
DATE             30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
                   15-Aug-1997
ACCESSIONS       JC1282
REFERENCE        JC1282
   #authors      Dai, J.; Baxter, R.C.
   #journal      Biochem. Biophys. Res. Commun. (1992) 188:304-309
   #title        Molecular cloning of the acid-labile subunit of the rat
                   insulin-like growth factor binding protein complex.
   #cross-references MUID:93038676
   #accession    JC1282
      ##molecule_type mRNA
      ##residues      1-603 ##label DAI
      ##experimental_source liver
      ##note          the authors translated the codon AAG for residue 63 as
                       Arg, AAA for residue 205 as Pro and GGT for residue
                       260 as Arg
CLASSIFICATION   #superfamily leucine-rich alpha-2-glycoprotein repeat
                   homology
FEATURE
   1-27              #domain signal sequence #status predicted #label SIG\
   28-603            #product insulin-like growth factor binding protein,
                       acid labile chain #status predicted #label MAT
SUMMARY          #length 603  #molecular-weight 66811  #checksum 8075

  Query Match             6.7%;  Score 104;  DB 2;  Length 603;
  Best Local Similarity  31.9%;  Pred. No. 1.51e+00;
  Matches    38;  Conservative   31;  Mismatches  36;  Indels  14;  Gaps  13;

Db    272 DLSHNRVAGLMEDTFPGLLGLHVLRLAHNAIASLRPRTFKDLHFLEELQLGHNRIRQLGE 331
          :|| :|:| |:  |   ||  : |||:  : |||:| :| :   |:|| : :: | :|:
Qy     76 NLSEERLAVLLAGT-HTLLQ-QALRLP--P-ASLKPDAFQE-E-LQELGIPQDLIGDLAS 128

Db    332 RTFEGLGQLEVL-TLNDNQITEV-RVGAFSGLFNVAVMNLSGNCLRSLPERVFQGLDKL 388
           :| | :|  :| :: : | : : :|: |    :||: : |:   |||   |:  | ||
Qy    129 LAF-G-SQRPLLDSVAQQQGSSLPHVSYFRWRVDVAI-STSAQS-RSLQPSVLMQL-KL 182


RESULT   10
ENTRY            G65036     #type complete
TITLE            hypothetical protein b2584 - Escherichia coli (strain K-12)
ORGANISM         #formal_name Escherichia coli
DATE             12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
                   14-Nov-1997
ACCESSIONS       G65036
REFERENCE        A64720
   #authors      Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
                   Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
                   Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
                   Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
                   Y.
   #journal      Science (1997) 277:1453-1462
   #title        The complete genome sequence of Escherichia coli K-12.
   #cross-references MUID:97426617
   #accession    G65036
      ##status       preliminary; nucleic acid sequence not shown;
                       translation not shown
      ##molecule_type DNA
      ##residues      1-886 ##label BLAT
      ##cross-references GB:AE000344; GB:U00096; NID:g1788927; PID:g1788938;
                        UWGP:b2584
      ##experimental_source strain K-12, substrain MG1655
SUMMARY          #length 886  #molecular-weight 97986  #checksum 3673

  Query Match             6.7%;  Score 104;  DB 2;  Length 886;
  Best Local Similarity  27.2%;  Pred. No. 1.51e+00;
  Matches    22;  Conservative   19;  Mismatches  39;  Indels   1;  Gaps   1;

Db    416 RLFSEAGLPTYRTPEGTITAFMHMVEYRRNQKQLRETPALPSN-LTSNTAEAHLLLQQAI 474
          :|: :   :|:|     : : :|  : |    ||  :: |: : |:   | :| |||||:
Qy     38 QLLKDLDRSTFRKLLKLVVGALHGKDCREAVEQLGASANLSEERLAVLLAGTHTLLQQAL 97

Db    475 AEGATSLDTHEVQPILQAYGM 495
            ::||     |  ||  |:
Qy     98 RLPPASLKPDAFQEELQELGI 118


RESULT   11
ENTRY            S77213     #type complete
TITLE            DNA-directed DNA polymerase (EC 2.7.7.7) III chain dnaX -
                   Synechocystis sp. (strain PCC 6803)
ALTERNATE_NAMES  protein sll1360
ORGANISM         #formal_name Synechocystis sp.
   #variety      PCC 6803
DATE             25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
                   21-Aug-1998
ACCESSIONS       S77213
REFERENCE        S74322
   #authors      Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
                   Nakamura, Y.; Miyajima, N.; Hirosawa, M.; Sugiura, M.;
                   Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
                   Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo,
                   S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
                   Yasuda, M.; Tabata, S.
   #journal      DNA Res. (1996) 3:109-136
   #title        Sequence analysis of the genome of the unicellular
                   cyanobacterium Synechocystis sp. PCC6803. II. Sequence
                   determination of the entire genome and assignment of
                   potential protein-coding regions.
   #cross-references MUID:97061201
   #accession    S77213
      ##status       nucleic acid sequence not shown; translation not shown
      ##molecule_type DNA
      ##residues      1-1116 ##label KAN
      ##cross-references EMBL:D90907; GB:AB001339; NID:g1652618; PID:d1018280;
                        PID:g1652627
      ##note          the nucleotide sequence was submitted to the EMBL Data
                       Library, June 1996
GENETICS
   #gene         dnaX
KEYWORDS         nucleotidyltransferase
SUMMARY          #length 1116  #molecular-weight 122009  #checksum 8010

  Query Match             6.7%;  Score 104;  DB 2;  Length 1116;
```

```
  Best Local Similarity  24.5%;  Pred. No. 1.51e+00;
  Matches    35;  Conservative   34;  Mismatches  67;  Indels   7;  Gaps   7;

Db    563 LSTAAFNALLKTLEEPPERVVFVFLATTDPQ-RVLPTIISRCQRFDYRRIPLQAMVDHLRY 621
          |::||   |    :  : || |: :   |:   :: ::   :|  :|:: |  :|  |:
Qy      4 LGAAA-PYLHHPADSHSGRVSFLGSQPSPEVTAVAQLLKDLDRSTFRKL-LKLVVGALHG 61

Db    622 IAGRENININDQPALTLVAQIANGGLRDAESLLDQ-LSLLPDLITPDKVWD-LVG-AVPEQ 678
            || ::   ::  |  :      |  : :||:| | | |  : ||   : |   ::| |
Qy     62 KDCREAVEQLGASANLSEERLAVLLAGTHTLLQQALRLPPASLKPDAFQEELQELGIP-Q 120

Db    679 DLLALLEAIASDDAETLLATCRQ 701
          ||:: | ::|      || :  |
Qy    121 DLIGDLASLAFGSQRPLLDSVAQ 143


RESULT   12
ENTRY            T05656     #type complete
TITLE            hypothetical protein F22I13.50 - Arabidopsis thaliana
ORGANISM         #formal_name Arabidopsis thaliana #common_name mouse-ear
                   cress
DATE             23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change
                   23-Jul-1999
ACCESSIONS       T05656
REFERENCE        Z15420
   #authors      Bevan, M.; Wedler, H.; Kutzner, M.; Wambutt, R.; Bancroft,
                   I.; Mewes, H.W.; Mayer, K.F.X.; Schueller, C.
   #submission   submitted to the Protein Sequence Database, February 1999
   #accession    T05656
      ##molecule_type DNA
      ##residues      1-159 ##label BEV
      ##cross-references EMBL:AL035539
      ##experimental_source cultivar Columbia; BAC clone F22I13
GENETICS
   #map_position 4
   #introns      117/1
   #note         F22I13.50
SUMMARY          #length 159  #molecular-weight 17633  #checksum 5499

  Query Match             6.6%;  Score 102;  DB 2;  Length 159;
  Best Local Similarity  25.0%;  Pred. No. 2.52e+00;
  Matches    23;  Conservative   30;  Mismatches  32;  Indels   7;  Gaps   7;

Db     22 SIGAKKPPLESPATTNAAS-GRLVYVRRRVEVDTSKAAASTTNPNPPPTKAPLQIPSSPA 80
          ::|: :| |:| |   ::|  :: | |:||:|  | :| | :   |:  :  | :  ::|
Qy    130 AFGSQRPLLDSVAQQQGSSLPHVSYFRWRVDVAISTSAQSRS-LQPSVLM-QLKLTDGSA 187

Db     81 QE-PTPTSHKLDWEERYLHLQMLLNKLNQSDR 111
          :    | : |:: | ||  : ::|: : : ::
Qy    188 HRFEVPIA-KFQ-ELRY-SVALVLKEMAELEK 216


RESULT   13
ENTRY            S23159     #type complete
TITLE            nucleoprotein N - Impatiens necrotic spot virus
ORGANISM         #formal_name Impatiens necrotic spot virus
DATE             22-Nov-1993 #sequence_revision 01-Dec-1995 #text_change
                   20-Sep-1999
ACCESSIONS       S23159
REFERENCE        S23158
   #authors      de Haan, P.; de Avila, A.C.; Kormelink, R.; Westerbroek, A.;
                   Gielen, J.J.L.; Peters, D.; Goldbach, R.
   #journal      FEBS Lett. (1992) 306:27-32
   #title        The nucleotide sequence of the S RNA of Impatiens necrotic
                   spot virus, a novel tospovirus.
   #cross-references MUID:92331780
   #accession    S23159
      ##status        preliminary
      ##molecule_type nucleic acid
      ##residues      1-262 ##label HAA
      ##cross-references EMBL:X66972; NID:g60488; PIDN:CAA47383.1; PID:g60490
CLASSIFICATION   #superfamily tomato spotted wilt virus nucleoprotein N
SUMMARY          #length 262  #molecular-weight 28718  #checksum 7578

  Query Match             6.6%;  Score 103;  DB 2;  Length 262;
  Best Local Similarity  36.2%;  Pred. No. 1.96e+00;
  Matches    17;  Conservative   12;  Mismatches  17;  Indels   1;  Gaps   1;

Db    137 ATTDLKSCVLGVLLGGSLPLIASVLNFEIAALVPAIYQDAKHVELGI 183
          |:::|    |:|||:|:  |:   | :  |:| |  :|:  : ||||
Qy     73 ASANLSEERLAVLLAGTHTLLQQALRLPPASLKPDAFQEELQ-ELGI 118


RESULT   14
ENTRY            S74728     #type complete
TITLE            hypothetical protein slr0971 ccmA 3'-region - Synechocystis
                   sp. (strain PCC 6803)
ORGANISM         #formal_name Synechocystis sp.
   #variety      PCC 6803
DATE             25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
                   21-Aug-1998
ACCESSIONS       S74728; B55214
REFERENCE        S74322
   #authors      Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
                   Nakamura, Y.; Miyajima, N.; Hirosawa, M.; Sugiura, M.;
                   Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
                   Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo,
                   S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
                   Yasuda, M.; Tabata, S.
   #journal      DNA Res. (1996) 3:109-136
   #title        Sequence analysis of the genome of the unicellular
                   cyanobacterium Synechocystis sp. PCC6803. II. Sequence
                   determination of the entire genome and assignment of
                   potential protein-coding regions.
   #cross-references MUID:97061201
   #accession    S74728
      ##status        nucleic acid sequence not shown; translation not shown
      ##molecule_type DNA
      ##residues      1-451 ##label KAN
      ##cross-references EMBL:D90901; GB:AB001339; NID:g1651897; PID:d1017612;
                        PID:g1651953
      ##note          the nucleotide sequence was submitted to the EMBL Data
                       Library, June 1996
REFERENCE        A55214
   #authors      Ogawa, T.; Marco, E.; Orus, M.I.
   #journal      J. Bacteriol. (1994) 176:2374-2378
   #title        A gene (ccmA) required for carboxysome formation in the
                   cyanobacterium Synechocystis sp. strain PCC6803.
   #cross-references MUID:94209239
   #accession    B55214
      ##molecule_type DNA
      ##residues      'VDPNLEVAQISNRMLQAGGPGLLFENVKG',
                       'SPFPVAVNLMGTVERICWAMNMDHPLELEDLGKK',244-451 ##label
                       OGA
      ##cross-references GB:D26444; NID:g452237; PID:d1006006; PID:g499299
      ##note          the authors translated the codon CCC for residue 236 as
                       Glu; the authors translated used GTG as an initiation
                       codon and translated it as Val
      ##note          the authors say that codon usages suggest this protein
                       is not expressed at high levels
SUMMARY          #length 451  #molecular-weight 49667  #checksum 7763

  Query Match             6.6%;  Score 102;  DB 2;  Length 451;
  Best Local Similarity  45.5%;  Pred. No. 2.52e+00;
  Matches    25;  Conservative    5;  Mismatches  21;  Indels   4;  Gaps   4;

Db     42 ALHFDDPRLPEVEQLTLSGT-PEE-LEALALAVETYLQQRLN-PCASLEPSTYPE 93
          |||  | |   ||||  |:  :|| |  |  :  | ||| |  | ||| | :: |
Qy     58 ALHGKDCR-EAVEQLGASANLSEERLAVLLAGTHTLLQQALRLPPASLKPDAFQE 111


RESULT   15
ENTRY            F72111     #type complete
TITLE            fructose-6-p phosphotransferase - Chlamydia pneumoniae
                   (strain CWL029)
ORGANISM         #formal_name Chlamydia pneumoniae
```

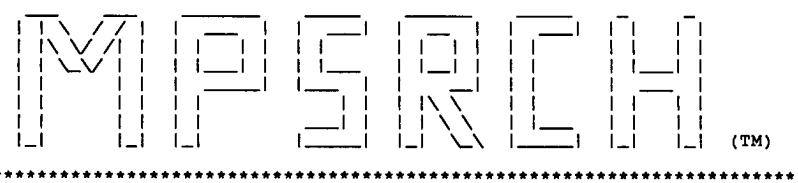
```
DATE              23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change
                    20-Sep-1999
ACCESSIONS        F72111
REFERENCE         A72000
   #authors       Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.;
                    Olinger, L.; Grimwood, J.; Davis, R.W.; Stephens, R.S.
   #journal       Nature Genet. (1999) 21:385-389
   #title         Comparative genomes of Clamydia pneumoniae and C.
                    trachomatis.
   #cross-references MUID:99206606
   #accession     F72111
      ##status         preliminary
      ##molecule_type DNA
      ##residues       1-555 ##label ARN
      ##cross-references GB:AE001603; GB:AE001363; NID:g4376422; PID:g4376428
      ##experimental_source strain CWL029
GENETICS
   #gene          pfkA_1
CLASSIFICATION    #superfamily pyrophosphate--fructose-6-phosphate
                    1-phosphotransferase alpha chain; 6-phosphofructokinase 1
                    homology
SUMMARY           #length 555  #molecular-weight 62414  #checksum 6064

  Query Match             6.6%;  Score 103;  DB 2;  Length 555;
  Best Local Similarity  23.1%;  Pred. No. 1.96e+00;
  Matches    21;  Conservative   29;  Mismatches  35;  Indels   6;  Gaps   6;

Db    315 IEHIFDTRKLIDE-LNVLLANGDSSIEKILSKLSPETLKTFHLFPKDIANQLLLARDSHG 373
          :|::  : :| :| | ||||   : ::  | :|:| :||    |  ::  :| :::|  |
Qy     68 VEQLGASANLSEERLAVLLAGTHTLLQQAL-RLPPASLKP-DAFQEEL-QELGIPQDLIG 124

Db    374 NVRVSKIATE-ELLAVMVKKEIEKIKPHMEF 403
          ::    ::::  ||  :   :   : ||: :
Qy    125 DLASLAFGSQRPLLDSVAQQQGSSL-PHVSY 154


Search completed: Fri May 12 10:26:57 2000
Job time : 58 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 12 10:28:07 2000; MasPar time 17.96 Seconds
161.600 Million cell updates/sec
Tabular output not generated.

Title: >US-09-223-796-2
Description: (1-224) from US09223796.pep
Perfect Score: 1552
Sequence: 1 MSALGAAAPYLHHPADSHSG.....ALVLKEMAELEKKCERKLQD 224

Scoring table: PAM 150
Gap 11

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT_COMB 4:backfiles1

Statistics: Mean 30.351; Variance 153.535; scale 0.198

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	124	8.0	195	2	US-08-822-	Sequence 3, Applicatio	2.71e-01
2	118	7.6	195	2	US-08-822-	Sequence 1, Applicatio	7.00e-01
3	104	6.7	603	1	US-08-190-	Sequence 50, Applicati	6.12e+00
4	98	6.3	605	1	US-08-190-	Sequence 49, Applicati	1.51e+01
5	95	6.1	912	1	US-07-789-	Sequence 8, Applicatio	2.37e+01
6	95	6.1	912	1	US-08-487-	Sequence 8, Applicatio	2.37e+01
7	95	6.1	912	1	US-08-005-	Sequence 8, Applicatio	2.37e+01
8	94	6.1	4472	2	US-08-804-	Sequence 2, Applicatio	2.75e+01
9	93	6.0	405	3	PCT-US93-1	Sequence 2, Applicatio	3.18e+01
10	92	5.9	681	2	US-08-346-	Sequence 37, Applicati	3.69e+01
11	92	5.9	681	1	US-08-083-	Sequence 18, Applicati	3.69e+01
12	92	5.9	1078	1	US-08-083-	Sequence 11, Applicati	3.69e+01
13	92	5.9	1078	2	US-08-346-	Sequence 32, Applicati	3.69e+01
14	92	5.9	1078	1	US-08-465-	Sequence 32, Applicati	3.69e+01
15	92	5.9	1078	1	US-08-264-	Sequence 32, Applicati	3.69e+01
16	92	5.9	2544	2	US-08-576-	Sequence 32, Applicati	3.69e+01
17	92	5.9	2556	1	US-08-083-	Sequence 20, Applicati	3.69e+01
18	88	5.7	301	1	US-08-420-	Sequence 47, Applicati	6.61e+01
19	88	5.7	301	2	US-08-343-	Sequence 22, Applicati	6.61e+01
20	89	5.7	1114	2	US-08-576-	Sequence 31, Applicati	5.71e+01
21	87	5.6	122	2	US-08-475-	Sequence 19, Applicati	7.63e+01
22	87	5.6	122	1	US-08-476-	Sequence 19, Applicati	7.63e+01
23	87	5.6	122	1	US-07-956-	Sequence 19, Applicati	7.63e+01
24	87	5.6	122	1	US-08-485-	Sequence 19, Applicati	7.63e+01
25	87	5.6	2396	1	US-08-157-	Sequence 2, Applicatio	7.63e+01
26	85	5.5	365	2	US-08-481-	Sequence 10, Applicati	1.02e+02
27	86	5.5	368	1	US-08-423-	Sequence 35, Applicati	8.81e+01
28	85	5.5	379	1	US-08-484-	Sequence 2, Applicatio	1.02e+02
29	85	5.5	379	1	US-08-347-	Sequence 2, Applicatio	1.02e+02
30	85	5.5	379	3	PCT-US95-0	Sequence 4, Applicatio	1.02e+02
31	85	5.5	379	3	PCT-US95-0	Sequence 2, Applicatio	1.02e+02
32	85	5.5	379	3	PCT-US94-0	Sequence 2, Applicatio	1.02e+02
33	85	5.5	379	3	PCT-US95-1	Sequence 2, Applicatio	1.02e+02
34	86	5.5	393	1	US-08-423-	Sequence 33, Applicati	8.81e+01
35	86	5.5	393	1	US-08-530-	Sequence 11, Applicati	8.81e+01
36	85	5.5	560	1	US-07-683-	Sequence 1, Applicatio	1.02e+02
37	86	5.5	746	3	PCT-US95-1	Sequence 2, Applicatio	8.81e+01
38	86	5.5	2254	2	US-08-677-	Sequence 3, Applicatio	8.81e+01
39	86	5.5	2254	2	US-08-790-	Sequence 3, Applicatio	8.81e+01
40	86	5.5	2482	1	US-08-328-	Sequence 6, Applicatio	8.81e+01
41	85	5.5	2509	1	US-08-469-	Sequence 10, Applicati	1.02e+02
42	85	5.5	2556	1	US-08-185-	Sequence 17, Applicati	1.02e+02
43	86	5.5	3248	1	US-08-353-	Sequence 1, Applicatio	8.81e+01
44	86	5.5	3248	3	PCT-US95-1	Sequence 1, Applicatio	8.81e+01
45	85	5.5	15281	2	US-08-471-	Sequence 2, Applicatio	1.02e+02

ALIGNMENTS

```
RESULT    1
ID   US-08-822-260-3     STANDARD;      PRT;   195 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 3, Application US/08822260
XX
CC    Sequence 3, Application US/08822260
CC    Patent No. 5830660
CC     GENERAL INFORMATION:
CC       APPLICANT:  Hillman, Jennifer L.
CC       APPLICANT:  Goli, Surya K.
CC       TITLE OF INVENTION:  NOVEL TUMORIGENESIS PROTEIN
CC       NUMBER OF SEQUENCES:  3
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Incyte Pharmaceuticals, Inc.
CC         STREET:  3174 Porter Drive
CC         CITY:  Palo Alto
CC         STATE:  CA
CC         COUNTRY:  USA
CC         ZIP:  94304
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette
CC         COMPUTER:  IBM Compatible
CC         OPERATING SYSTEM:  DOS
CC         SOFTWARE:  FastSEQ for Windows Version 2.0
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/822,260
CC         FILING DATE:  Herewith
CC         CLASSIFICATION:  514
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:
CC         FILING DATE:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Billings, Lucy J.
CC         REGISTRATION NUMBER:  36,749
CC         REFERENCE/DOCKET NUMBER:  PF-0247 US
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  415-855-0555
CC         TELEFAX:  415-845-4166
CC         TELEX:
CC     INFORMATION FOR SEQ ID NO:  3:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  195 amino acids
CC         TYPE:  amino acid
```

```
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       IMMEDIATE SOURCE:
CC         LIBRARY:  GenBank
CC         CLONE:  265569
SQ   SEQUENCE   195 AA; 22037 MW; 190094 CN;

  Query Match            8.0%;  Score 124;  DB 2;  Length 195;
  Best Local Similarity  22.1%;  Pred. No. 2.71e-01;
  Matches   21;  Conservative   33;  Mismatches  37;  Indels   4;  Gaps   4;

Db     98 ELFCTEYQNNKNSLETLLGSIGRSLPHITDVSWRLEYQIKTNQLHKMYRPGYLVTLNVEN 157
           :|    : : :  |:::    | ||||::   ||::  | |:   :  :|: |: |:: :
Qy    125 DLASLAFGSQRPLLDSVAQQQGSSLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQLKLTD 184

Db    158 NDSQSYPEINFSCNMEQLQDLVG-KLKDASKSLER 191
            :: : |: :: : ::|:  |:  ||: :  ||:
Qy    185 GSAHRF-EVPIA-KFQELRYSVALVLKEMAE-LEK 216


RESULT    2
ID   US-08-822-260-1     STANDARD;     PRT;   195 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 1, Application US/08822260
XX
CC    Sequence 1, Application US/08822260
CC    Patent No. 5830660
CC     GENERAL INFORMATION:
CC       APPLICANT:  Hillman, Jennifer L.
CC       APPLICANT:  Goli, Surya K.
CC       TITLE OF INVENTION:  NOVEL TUMORIGENESIS PROTEIN
CC       NUMBER OF SEQUENCES:  3
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Incyte Pharmaceuticals, Inc.
CC         STREET:  3174 Porter Drive
CC         CITY:  Palo Alto
CC         STATE:  CA
CC         COUNTRY:  USA
CC         ZIP:  94304
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette
CC         COMPUTER:  IBM Compatible
CC         OPERATING SYSTEM:  DOS
CC         SOFTWARE:  FastSEQ for Windows Version 2.0
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/822,260
CC         FILING DATE:  Herewith
CC         CLASSIFICATION:  514
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:
CC         FILING DATE:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Billings, Lucy J.
CC         REGISTRATION NUMBER:  36,749
CC         REFERENCE/DOCKET NUMBER:  PF-0247 US
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  415-855-0555
CC         TELEFAX:  415-845-4166
CC         TELEX:
CC      INFORMATION FOR SEQ ID NO:  1:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  195 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       IMMEDIATE SOURCE:
CC         LIBRARY:  UTRSNOT02
CC         CLONE:  2267574
SQ   SEQUENCE   195 AA; 22151 MW; 196728 CN;

  Query Match            7.6%;  Score 118;  DB 2;  Length 195;
  Best Local Similarity  23.2%;  Pred. No. 7.00e-01;
  Matches   22;  Conservative   29;  Mismatches  40;  Indels   4;  Gaps   4;

Db     98 ELFCTEYQNNKNSLEILLGSIGRSLPHITDVSWRLEYQIKTNQLHRMYRPAYLVTLSVQN 157
           :|    : : :  |: :    | ||||::   ||::  | |:   |  :|: ||: | : :
Qy    125 DLASLAFGSQRPLLDSVAQQQGSSLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQLKLTD 184

Db    158 TDSPSYPEISFSCSMEQLQDLVG-KLKDASKSLER 191
            :  : |::::   ::|:  |:  ||: :  ||:
Qy    185 GSAHRF-EVPIA-KFQELRYSVALVLKEMAE-LEK 216


RESULT    3
ID   US-08-190-802A-50     STANDARD;     PRT;   603 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 50, Application US/08190802A
XX
CC    Sequence 50, Application US/08190802A
CC    Patent No. 5519003
CC     GENERAL INFORMATION:
CC       APPLICANT:  Mochly-Rosen, Daria
CC       APPLICANT:  Ron, Dorit
CC       TITLE OF INVENTION:  WD-40 - Derived Peptides and Uses
CC       TITLE OF INVENTION:  Thereof
CC       NUMBER OF SEQUENCES:  265
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Dehlinger & Associates
CC         STREET:  P.O. Box 60850
CC         CITY:  Palo Alto
CC         STATE:  CA
CC         COUNTRY:  USA
CC         ZIP:  94306-0850
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/190,802A
CC         FILING DATE:  01-FEB-1994
CC         CLASSIFICATION:  530
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Fabian, Gary R.
CC         REGISTRATION NUMBER:  33,875
CC         REFERENCE/DOCKET NUMBER:  8600-0139
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (415) 324-0880
CC         TELEFAX:  (415) 324-0960
CC      INFORMATION FOR SEQ ID NO:  50:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  603 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  unknown
CC       MOLECULE TYPE:  protein
CC       HYPOTHETICAL:  NO
CC       ANTI-SENSE:  NO
CC       ORIGINAL SOURCE:
CC         INDIVIDUAL ISOLATE:  Insulin-like growth factor bind
CC         INDIVIDUAL ISOLATE:  pro. complex-rat, Fig. 33
SQ   SEQUENCE   603 AA; 66811 MW; 1806049 CN;

  Query Match            6.7%;  Score 104;  DB 1;  Length 603;
  Best Local Similarity  31.9%;  Pred. No. 6.12e+00;
  Matches   38;  Conservative   31;  Mismatches  36;  Indels  14;  Gaps  13;

Db    272 DLSHNRVAGLMEDTFPGLLGLHVLRLAHNAIASLRPRTFKDLHFLEELQLGHNRIRQLGE 331
          :|| :|:| |:  |   ||  : |||:  : |||:| :| :   |:|| : :: | :|:
```

```
Qy      76 NLSEERLAVLLAGT-HTLLQ-QALRLP--P-ASLKPDAFQE-E-LQELGIPQDLIGDLAS 128

Db     332 RTFEGLGQLEVL-TLNDNQITEV-RVGAFSGLFNVAVMNLSGNCLRSLPERVFQGLDKL 388
           :| | :|  :| :: : | : : :|: |    :||: : |:   |||   |:  | ||
Qy     129 LAF-G-SQRPLLDSVAQQQGSSLPHVSYFRWRVDVAI-STSAQS-RSLQPSVLMQL-KL 182


RESULT    4
ID   US-08-190-802A-49      STANDARD;      PRT;   605 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 49, Application US/08190802A
XX
CC    Sequence 49, Application US/08190802A
CC    Patent No. 5519003
CC     GENERAL INFORMATION:
CC       APPLICANT:  Mochly-Rosen, Daria
CC       APPLICANT:  Ron, Dorit
CC       TITLE OF INVENTION:  WD-40 - Derived Peptides and Uses
CC       TITLE OF INVENTION:  Thereof
CC       NUMBER OF SEQUENCES:  265
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Dehlinger & Associates
CC         STREET:  P.O. Box 60850
CC         CITY:  Palo Alto
CC         STATE:  CA
CC         COUNTRY:  USA
CC         ZIP:  94306-0850
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/190,802A
CC         FILING DATE:  01-FEB-1994
CC         CLASSIFICATION:  530
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Fabian, Gary R.
CC         REGISTRATION NUMBER:  33,875
CC         REFERENCE/DOCKET NUMBER:  8600-0139
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (415) 324-0880
CC         TELEFAX:  (415) 324-0960
CC      INFORMATION FOR SEQ ID NO:  49:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  605 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  unknown
CC       MOLECULE TYPE:  protein
CC       HYPOTHETICAL:  NO
CC       ANTI-SENSE:  NO
CC       ORIGINAL SOURCE:
CC         INDIVIDUAL ISOLATE:  Insulin-like growth factor binding
CC         INDIVIDUAL ISOLATE:  protein complex, Fig. 32
SQ   SEQUENCE   605 AA; 66034 MW; 1699376 CN;

  Query Match             6.3%;  Score 98;  DB 1;  Length 605;
  Best Local Similarity  38.1%;  Pred. No. 1.51e+01;
  Matches    24;  Conservative   15;  Mismatches  17;  Indels   7;  Gaps   6;

Db     272 DLSHNRVAGLLEDTFPGLLGLRVLRLSHNAIASLRPRTFKDLHFLEELQLGHNRIRQLAE 331
           :|| :|:| ||  |   ||  : |||:  : |||:| :| :   |:|| : :: | :||
Qy      76 NLSEERLAVLLAGT-HTLLQ-QALRLP--P-ASLKPDAFQE-E-LQELGIPQDLIGDLAS 128

Db     332 RSF 334
           :|
Qy     129 LAF 131


RESULT    5
ID   US-07-789-915A-8      STANDARD;      PRT;   912 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 8, Application US/07789915A
XX
CC    Sequence 8, Application US/07789915A
CC    Patent No. 5212058
CC     GENERAL INFORMATION:
CC       APPLICANT:  Baker, Rohan T.
CC       APPLICANT:  Tobias, John W.
CC       APPLICANT:  Varshavsky, Alexander
CC       TITLE OF INVENTION:  Ubiquitin-Specific Proteases
CC       NUMBER OF SEQUENCES:  8
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
CC         STREET:  Two Militia Drive
CC         CITY:  Lexington
CC         STATE:  Massachusetts
CC         COUNTRY:  U.S.A.
CC         ZIP:  02173
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/07/789,915A
CC         FILING DATE:  19911108
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Brook, David E.
CC         REGISTRATION NUMBER:  22,592
CC         REFERENCE/DOCKET NUMBER:  MIT-5091AA
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  617-861-6240
CC         TELEFAX:  617-861-9540
CC      INFORMATION FOR SEQ ID NO:  8:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  912 amino acids
CC         TYPE:  AMINO ACID
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   912 AA; 101916 MW; 4327836 CN;

  Query Match             6.1%;  Score 95;  DB 1;  Length 912;
  Best Local Similarity  31.1%;  Pred. No. 2.37e+01;
  Matches    14;  Conservative   16;  Mismatches  15;  Indels   0;  Gaps   0;

Db     705 FRSVLDIPNNKESQSITLDPFQTIQLDISDAGVNDLETAFKKFSE 749
           ||  :|:: :  :|| :|:|   :|| ::|:: : :| :: || |
Qy     155 FRWRVDVAISTSAQSRSLQPSVLMQLKLTDGSAHRFEVPIAKFQE 199


RESULT    6
ID   US-08-487-203A-8      STANDARD;      PRT;   912 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 8, Application US/08487203A
XX
CC    Sequence 8, Application US/08487203A
CC    Patent No. 5683904
CC     GENERAL INFORMATION:
CC       APPLICANT:  Baker, Rohan T.
CC       APPLICANT:  Tobias, John W.
CC       APPLICANT:  Varshavsky, Alexander
CC       TITLE OF INVENTION:  Ubiquitin-Specific Proteases
```

```
CC       NUMBER OF SEQUENCES:  9
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Kevin M. Farrell, P.C.
CC         STREET:  P.O. Box 999
CC         CITY:  York Harbor
CC         STATE:  Maine
CC         COUNTRY:  U.S.A.
CC         ZIP:  03911
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/487,203A
CC         FILING DATE:  07-JUN-1995
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 08/005,002
CC         FILING DATE:  15-JAN-1993
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Farrell, Kevin M.
CC         REGISTRATION NUMBER:  35,505
CC         REFERENCE/DOCKET NUMBER:  MIT-5091A3Z
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  207-363-0558
CC         TELEFAX:  207-363-0528
CC     INFORMATION FOR SEQ ID NO:  8:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  912 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   912 AA; 101916 MW; 4327836 CN;

  Query Match             6.1%;  Score 95;  DB 1;  Length 912;
  Best Local Similarity  31.1%;  Pred. No. 2.37e+01;
  Matches    14;  Conservative   16;  Mismatches  15;  Indels   0;  Gaps   0;

Db    705 FRSVLDIPNNKESQSITLDPFQTIQLDISDAGVNDLETAFKKFSE 749
          ||  :|:: :  :|| :|:|   :|| ::|:: : :| :: || |
Qy    155 FRWRVDVAISTSAQSRSLQPSVLMQLKLTDGSAHRFEVPIAKFQE 199


RESULT    7
ID   US-08-005-002C-8     STANDARD;     PRT;   912 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 8, Application US/08005002C
XX
CC    Sequence 8, Application US/08005002C
CC    Patent No. 5494818
CC     GENERAL INFORMATION:
CC       APPLICANT:  Baker, Rohan T.
CC       APPLICANT:  Tobias, John W.
CC       APPLICANT:  Varshavsky, Alexander
CC       TITLE OF INVENTION:  Ubiquitin-Specific Proteases
CC       NUMBER OF SEQUENCES:  9
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Kevin M. Farrell, P.C.
CC         STREET:  P.O. Box 999
CC         CITY:  York Harbor
CC         STATE:  Maine
CC         COUNTRY:  U.S.A.
CC         ZIP:  03911
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/005,002C
CC         FILING DATE:  15-JAN-1993
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/789,915
CC         FILING DATE:  08-NOV-1991
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Farrell, Kevin M.
CC         REGISTRATION NUMBER:  35,505
CC         REFERENCE/DOCKET NUMBER:  MIT-5091AAZ
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  207-363-0558
CC         TELEFAX:  207-363-0528
CC     INFORMATION FOR SEQ ID NO:  8:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  912 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   912 AA; 101916 MW; 4327836 CN;

  Query Match             6.1%;  Score 95;  DB 1;  Length 912;
  Best Local Similarity  31.1%;  Pred. No. 2.37e+01;
  Matches    14;  Conservative   16;  Mismatches  15;  Indels   0;  Gaps   0;

Db    705 FRSVLDIPNNKESQSITLDPFQTIQLDISDAGVNDLETAFKKFSE 749
          ||  :|:: :  :|| :|:|   :|| ::|:: : :| :: || |
Qy    155 FRWRVDVAISTSAQSRSLQPSVLMQLKLTDGSAHRFEVPIAKFQE 199


RESULT    8
ID   US-08-804-227C-2     STANDARD;     PRT;   4472 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application US/08804227C
XX
CC    Sequence 2, Application US/08804227C
CC    Patent No. 5876991
CC     GENERAL INFORMATION:
CC       APPLICANT:  DeHoff, Bradley S.
CC       APPLICANT:  Kuhstoss, Stuart A.
CC       APPLICANT:  Rosteck, Paul R., Jr.
CC       APPLICANT:  Sutton, Kimberly L.
CC       TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
CC       NUMBER OF SEQUENCES:  15
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  THOMAS G. PLANT  1501
CC         STREET:  LILLY CORPORATE CENTER
CC         CITY:  INDIANAPOLIS
CC         STATE:  IN
CC         COUNTRY:  USA
CC         ZIP:  46285
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM Compatible
CC         OPERATING SYSTEM:  MS-DOS
CC         SOFTWARE:  ASCI(DOS) Text only
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/804,227C
CC         FILING DATE:  February 21, 1997
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Plant, Thomas, G.
CC         REGISTRATION NUMBER:  35,784
CC         REFERENCE/DOCKET NUMBER:  X-8231
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  317-276-2459
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
```

```
CC         LENGTH:  4472 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  unknown
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   4472 AA; 463281 MW; 93208012 CN;

  Query Match             6.1%;  Score 94;  DB 2;  Length 4472;
  Best Local Similarity  27.4%;  Pred. No. 2.75e+01;
  Matches    43;  Conservative   42;  Mismatches  59;  Indels  13;  Gaps  10;

Db   4287 GGAPAPAGGEGTAERLAGLGEEARLRELVRLVRAEVSGVLGYSGPDAVEPGRPFKDLGFD 4346
           |: |:|    : |: |  | : : :| |::||   |::: |    :|||   :  :|: :
Qy     26 GSQPSPEVT-AVAQLLKDL-DRSTFRKLLKLV---VGALHGKDCREAVEQLGASANLSEE 80

Db   4347 SLTAVELRNRLGAATGLRLPTALVFDRPTSQAVAEYLAAELAGPRD-GGDTAAAAFEGLE 4405
           |: :    :     :|||| |  : :|  :|  | |  ||: |:|  || |: || :
Qy     81 RLAVLLAGTHTLLQQALRLPPA-SL-KP--DAFQEELQ-ELGIPQDLIGDLASLAFGSQR 135

Db   4406 ALAAAVGALAEDDLRR-DVLRRRLTELAAALTPQGRN 4441
           :|  :|:      | :   :|:|: ::| : ::|:|:
Qy    136 PLLDSVAQQQGSSLPHVSYFRWRV-DVAISTSAQSRS 171


RESULT    9
ID   PCT-US93-11404-2     STANDARD;     PRT;   405 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application PC/TUS9311404
XX
CC    Sequence 2, Application PC/TUS9311404
CC     GENERAL INFORMATION:
CC       APPLICANT:  THE BOARD OF TRUSTEES OF LELAND STANFORD
CC       APPLICANT:  JUNIOR UNIVERSITY
CC       TITLE OF INVENTION:  MUCOSAL VASCULAR ADDRESSIN, DNA AND
CC       TITLE OF INVENTION:  EXPRESSION
CC       NUMBER OF SEQUENCES:  2
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Bertram I. Rowland
CC         STREET:  4 Embarcadero Center
CC         CITY:  San Francisco
CC         STATE:  California
CC         COUNTRY:  USA
CC         ZIP:  94111
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/US93/11404
CC         FILING DATE:  23-Nov-1993
CC         CLASSIFICATION:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Rowland, Bertram I.
CC         REGISTRATION NUMBER:  20,015
CC         REFERENCE/DOCKET NUMBER:  FP-57452/BIR;STAN-144
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (415) 781-1989
CC         TELEFAX:  (415) 398-3249
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  405 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   405 AA; 43651 MW; 918196 CN;

  Query Match             6.0%;  Score 93;  DB 3;  Length 405;
  Best Local Similarity  31.6%;  Pred. No. 3.18e+01;
  Matches    18;  Conservative   20;  Mismatches  13;  Indels   6;  Gaps   5;

Db    168 EEIQEAEGTPLFRMTQ-RWRLP-SLGTPAPP-ALHCQVTMQLPKLV--LTHRKEIPV 219
          : : : :|::| :::  |||:  :::|:| : :|:  | ||| ||    :|| |:|:
Qy    139 DSVAQQQGSSLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQL-KLTDGSAHRFEVPI 194


RESULT   10
ID   US-08-346-128-37     STANDARD;     PRT;   681 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 37, Application US/08346128
XX
CC    Sequence 37, Application US/08346128
CC    Patent No. 5856441
CC     GENERAL INFORMATION:
CC       APPLICANT:  Artavanis-Tsakonas, Spyridon et al.
CC       TITLE OF INVENTION:  Human No. 5856441ch And Delta, Binding Domains
CC       TITLE OF INVENTION:  In Toporythmic Proteins, And Methods Based Thereon
CC       NUMBER OF SEQUENCES:  37
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Pennie & Edmonds
CC         STREET:  1155 Avenue of the Americas
CC         CITY:  New York
CC         STATE:  New York
CC         COUNTRY:  U.S.A.
CC         ZIP:  10036
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/346,128
CC         FILING DATE:
CC         CLASSIFICATION:  530
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  07/879,038
CC         FILING DATE:  30-APR-1992
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Misrock, S. Leslie
CC         REGISTRATION NUMBER:  18,872
CC         REFERENCE/DOCKET NUMBER:  7326-009
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  212 790-9090
CC         TELEFAX:  212 8698864/9741
CC         TELEX:  66141 PENNIE
CC     INFORMATION FOR SEQ ID NO:  37:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  681 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  unknown
CC       MOLECULE TYPE:  peptide
SQ   SEQUENCE   681 AA; 71840 MW; 2534721 CN;

  Query Match             5.9%;  Score 92;  DB 2;  Length 681;
  Best Local Similarity  29.9%;  Pred. No. 3.69e+01;
  Matches    20;  Conservative   15;  Mismatches  29;  Indels   3;  Gaps   2;

Db    486 MVGPLHSSLAASALSQMMSYQGLPSTRLATQPHLVQTQQVQPQNLQMQQQNLQPANIQQQ 545
          :||:||:     |: |: :   |:  |||    |  |: :  | |::   :| |  :|::
Qy     55 VVGALHGKDCREAVEQLGASANLSEERLAVL--LAGTHTLLQQALRLPPASLKPDAFQEE 112

Db    546 -QSLQPP 551
           -| |  |
Qy    113 LQELGIP 119


RESULT   11
ID   US-08-083-590A-18     STANDARD;     PRT;   681 AA.
```

```
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 18, Application US/08083590A
XX
CC    Sequence 18, Application US/08083590A
CC    Patent No. 5786158
CC     GENERAL INFORMATION:
CC       APPLICANT:  Artavanis-Tsakonas, S. et al.
CC       TITLE OF INVENTION:  Therapeutic And Diagnostic Methods
CC       TITLE OF INVENTION:  And Compositions Based On No. 5786158ch Proteins And
CC       TITLE OF INVENTION:  Nucleic Acids
CC       NUMBER OF SEQUENCES:  21
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Pennie & Edmonds
CC         STREET:  1155 Avenue of the Americas
CC         CITY:  New York
CC         STATE:  New York
CC         COUNTRY:  U.S.A.
CC         ZIP:  10036
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/083,590A
CC         FILING DATE:  25-JUN-1993
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Misrock, S. Leslie
CC         REGISTRATION NUMBER:  18,872
CC         REFERENCE/DOCKET NUMBER:  7326-015
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  212 790-9090
CC         TELEFAX:  212 8698864/9741
CC         TELEX:  66141 PENNIE
CC     INFORMATION FOR SEQ ID NO:  18:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  681 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  unknown
CC       MOLECULE TYPE:  peptide
SQ   SEQUENCE   681 AA; 71840 MW; 2534721 CN;

  Query Match             5.9%;  Score 92;  DB 1;  Length 681;
  Best Local Similarity  29.9%;  Pred. No. 3.69e+01;
  Matches    20;  Conservative   15;  Mismatches  29;  Indels   3;  Gaps   2;

Db    486 MVGPLHSSLAASALSQMMSYQGLPSTRLATQPHLVQTQQVQPQNLQMQQQNLQPANIQQQ 545
          :||:||:     |: |: :   |:  |||    |  |: : | |::   :| |  :|::
Qy     55 VVGALHGKDCREAVEQLGASANLSEERLAVL--LAGTHTLLQQALRLPPASLKPDAFQEE 112

Db    546 -QSLQPP 551
           | |  |
Qy    113 LQELGIP 119


RESULT   12
ID   US-08-083-590A-11     STANDARD;     PRT;   1078 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 11, Application US/08083590A
XX
CC    Sequence 11, Application US/08083590A
CC    Patent No. 5786158
CC     GENERAL INFORMATION:
CC       APPLICANT:  Artavanis-Tsakonas, S. et al.
CC       TITLE OF INVENTION:  Therapeutic And Diagnostic Methods
CC       TITLE OF INVENTION:  And Compositions Based On No. 5786158ch Proteins And
CC       TITLE OF INVENTION:  Nucleic Acids
CC       NUMBER OF SEQUENCES:  21
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Pennie & Edmonds
CC         STREET:  1155 Avenue of the Americas
CC         CITY:  New York
CC         STATE:  New York
CC         COUNTRY:  U.S.A.
CC         ZIP:  10036
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/083,590A
CC         FILING DATE:  25-JUN-1993
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Misrock, S. Leslie
CC         REGISTRATION NUMBER:  18,872
CC         REFERENCE/DOCKET NUMBER:  7326-015
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  212 790-9090
CC         TELEFAX:  212 8698864/9741
CC         TELEX:  66141 PENNIE
CC     INFORMATION FOR SEQ ID NO:  11:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1078 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  unknown
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   1078 AA; 116254 MW; 5971068 CN;

  Query Match             5.9%;  Score 92;  DB 1;  Length 1078;
  Best Local Similarity  29.9%;  Pred. No. 3.69e+01;
  Matches    20;  Conservative   15;  Mismatches  29;  Indels   3;  Gaps   2;

Db    893 MVGPLHSSLAASALSQMMSYQGLPSTRLATQPHLVQTQQVQPQNLQMQQQNLQPANIQQQ 952
          :||:||:     |: |: :   |:  |||    |  |: : | |::   :| |  :|::
Qy     55 VVGALHGKDCREAVEQLGASANLSEERLAVL--LAGTHTLLQQALRLPPASLKPDAFQEE 112

Db    953 -QSLQPP 958
           | |  |
Qy    113 LQELGIP 119


RESULT   13
ID   US-08-346-128-32     STANDARD;     PRT;   1078 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 32, Application US/08346128
XX
CC    Sequence 32, Application US/08346128
CC    Patent No. 5856441
CC     GENERAL INFORMATION:
CC       APPLICANT:  Artavanis-Tsakonas, Spyridon et al.
CC       TITLE OF INVENTION:  Human No. 5856441ch And Delta, Binding Domains
CC       TITLE OF INVENTION:  In Toporythmic Proteins, And Methods Based Thereon
CC       NUMBER OF SEQUENCES:  37
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Pennie & Edmonds
CC         STREET:  1155 Avenue of the Americas
CC         CITY:  New York
CC         STATE:  New York
CC         COUNTRY:  U.S.A.
CC         ZIP:  10036
```

```
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/346,128
CC         FILING DATE:
CC         CLASSIFICATION:  530
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  07/879,038
CC         FILING DATE:  30-APR-1992
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Misrock, S. Leslie
CC         REGISTRATION NUMBER:  18,872
CC         REFERENCE/DOCKET NUMBER:  7326-009
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  212 790-9090
CC         TELEFAX:  212 8698864/9741
CC         TELEX:  66141 PENNIE
CC     INFORMATION FOR SEQ ID NO:  32:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1078 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  unknown
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   1078 AA; 116254 MW; 5971068 CN;

  Query Match             5.9%;  Score 92;  DB 2;  Length 1078;
  Best Local Similarity  29.9%;  Pred. No. 3.69e+01;
  Matches    20;  Conservative   15;  Mismatches  29;  Indels   3;  Gaps   2;

Db    893 MVGPLHSSLAASALSQMMSYQGLPSTRLATQPHLVQTQQVQPQNLQMQQQNLQPANIQQQ 952
          :||:||:     |: |: :   |:  |||    |  |: :  | |::   :| |  :|::
Qy     55 VVGALHGKDCREAVEQLGASANLSEERLAVL--LAGTHTLLQQALRLPPASLKPDAFQEE 112

Db    953 -QSLQPP 958
           | |  |
Qy    113 LQELGIP 119


RESULT   14
ID   US-08-465-500-32     STANDARD;     PRT;   1078 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 32, Application US/08465500
XX
CC    Sequence 32, Application US/08465500
CC    Patent No. 5789195
CC     GENERAL INFORMATION:
CC       APPLICANT:  Artavanis-Tsakonas, Spyridon
CC       APPLICANT:  Muskavitch, Marc A.T.
CC       APPLICANT:  Fehon, Richard G.
CC       APPLICANT:  Rebay, Ilaria
CC       APPLICANT:  Blaumueller, Cristine M.
CC       APPLICANT:  Shepard, Scott B.
CC       TITLE OF INVENTION:  HUMAN NOTCH AND DELTA, BINDING DOMAINS
CC       TITLE OF INVENTION:  IN TOPORYTHMIC PROTEINS,AND METHODS BASED THEREON
CC       NUMBER OF SEQUENCES:  34
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  PENNIE & EDMONDS
CC         STREET:  1155 Avenue of the Americas
CC         CITY:  New York
CC         STATE:  NY
CC         COUNTRY:  USA
CC         ZIP:  10036-2711
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/465,500
CC         FILING DATE:  05-JUN-1995
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Misrock, S. Leslie
CC         REGISTRATION NUMBER:  18,872
CC         REFERENCE/DOCKET NUMBER:  7326-034
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (212) 790-9090
CC         TELEFAX:  (212) 869-8864/9741
CC     INFORMATION FOR SEQ ID NO:  32:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1078 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  unknown
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   1078 AA; 116254 MW; 5971068 CN;

  Query Match             5.9%;  Score 92;  DB 1;  Length 1078;
  Best Local Similarity  29.9%;  Pred. No. 3.69e+01;
  Matches    20;  Conservative   15;  Mismatches  29;  Indels   3;  Gaps   2;

Db    893 MVGPLHSSLAASALSQMMSYQGLPSTRLATQPHLVQTQQVQPQNLQMQQQNLQPANIQQQ 952
          :||:||:     |: |: :   |:  |||    |  |: :  | |::   :| |  :|::
Qy     55 VVGALHGKDCREAVEQLGASANLSEERLAVL--LAGTHTLLQQALRLPPASLKPDAFQEE 112

Db    953 -QSLQPP 958
           | |  |
Qy    113 LQELGIP 119


RESULT   15
ID   US-08-264-534-32     STANDARD;     PRT;   1078 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 32, Application US/08264534
XX
CC    Sequence 32, Application US/08264534
CC    Patent No. 5648464
CC     GENERAL INFORMATION:
CC       APPLICANT:  Artavanis-Tsakonas, Spyridon et al.
CC       TITLE OF INVENTION:  Human No. 5648464ch And Delta, Binding Domains
CC       TITLE OF INVENTION:  In Toporythmic Proteins, And Methods Based Thereon
CC       NUMBER OF SEQUENCES:  34
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Pennie & Edmonds
CC         STREET:  1155 Avenue of the Americas
CC         CITY:  New York
CC         STATE:  New York
CC         COUNTRY:  U.S.A.
CC         ZIP:  10036
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/264,534
CC         FILING DATE:
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/695,189
CC         FILING DATE:  03-MAY-1991
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Misrock, S. Leslie
CC         REGISTRATION NUMBER:  18,872
CC         REFERENCE/DOCKET NUMBER:  7326-004
```

```
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  212 790-9090
CC         TELEFAX:  212 8698864/9741
CC         TELEX:  66141 PENNIE
CC     INFORMATION FOR SEQ ID NO:  32:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1078 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  unknown
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   1078 AA; 116254 MW; 5971068 CN;

  Query Match             5.9%;  Score 92;  DB 1;  Length 1078;
  Best Local Similarity  29.9%;  Pred. No. 3.69e+01;
  Matches    20;  Conservative   15;  Mismatches  29;  Indels   3;  Gaps   2;

Db    893 MVGPLHSSLAASALSQMMSYQGLPSTRLATQPHLVQTQQVQPQNLQMQQQNLQPANIQQQ 952
          :||:||:     |: |: :   |:  |||    |  |: :  | |::   :| |  :|::
Qy     55 VVGALHGKDCREAVEQLGASANLSEERLAVL--LAGTHTLLQQALRLPPASLKPDAFQEE 112

Db    953 -QSLQPP 958
           | |  |
Qy    113 LQELGIP 119


Search completed: Fri May 12 10:28:29 2000
Job time : 22 secs.
```

****

MPSRCH (TM)

****

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 12 10:27:14 2000; MasPar time 29.86 Seconds
177.710 Million cell updates/sec

Tabular output not generated.

Title: >US-09-223-796-2
Description: (1-224) from US09223796.pep
Perfect Score: 1552
Sequence: 1 MSALGAAAPYLHHPADSHSG.....ALVLKEMAELEKKCERKLQD 224

Scoring table: PAM 150
Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq36
1:geneseqp

Statistics: Mean 32.859; Variance 162.439; scale 0.202

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	1552	100.0	224	1	W37723	Rat Hypertension relat	2.90e-120
2	119	7.7	169	1	W88390	Human Zneu1 partial po	1.44e+00
3	119	7.7	181	1	W88391	Human Zneu1 partial po	1.44e+00
4	119	7.7	254	1	W88382	Human neuro-growth fac	1.44e+00
5	119	7.7	273	1	W88381	Human neuro-growth fac	1.44e+00
6	118	7.6	195	1	W71684	Amino acid sequence of	1.69e+00
7	118	7.6	196	1	Y07867	Human secreted protein	1.69e+00
8	104	6.7	603	1	R85889	WD-40 domain-contg. ra	1.46e+01
9	103	6.6	262	1	R42079	Impatiens Necrotic Spo	1.69e+01
10	98	6.3	605	1	R85888	WD-40 domain-contg. in	3.57e+01
11	97	6.2	178	1	W95014	Chicken MGF sequence.	4.14e+01
12	95	6.1	912	1	R36731	Ubiquitin-specific pro	5.56e+01
13	94	6.1	4472	1	W22601	Tylactone synthase ORF	6.43e+01
14	93	6.0	166	1	W05049	Human thrombopoietin d	7.44e+01
15	93	6.0	405	1	R57140	Mouse mucosal addressi	7.44e+01
16	93	6.0	405	1	W60615	Mouse mucosal adressin	7.44e+01
17	93	6.0	547	1	R26189	Epidemic NANBH virus a	7.44e+01
18	91	5.9	193	1	W14142	Erythropoietin variant	9.95e+01
19	91	5.9	390	1	W97585	An ACI-1 polypeptide.	9.95e+01
20	91	5.9	920	1	W82500	Human OGT protein.	9.95e+01
21	91	5.9	1078	1	R28963	Notch hN3k full length	9.95e+01
22	90	5.8	796	1	R84618	Oilseed rape acetyl Co	1.15e+02
23	89	5.7	332	1	R99525	Neutral protease of Cl	1.33e+02
24	88	5.7	332	1	W12929	Mpl ligand analogue, [	1.53e+02
25	88	5.7	350	1	W60244	Amino acid of the spec	1.53e+02
26	89	5.7	360	1	W19734	Sugar biosynthesis enz	1.33e+02
27	89	5.7	361	1	W99387	S.erythraea dTDP-4-ket	1.33e+02
28	88	5.7	440	1	W42391	Thermococcus celer pho	1.53e+02
29	89	5.7	567	1	R12517	Maize nitrite reductas	1.33e+02
30	89	5.7	936	1	W06719	Rhodopseudomonas sp. p	1.33e+02
31	88	5.7	1874	1	W64518	Adenylate cyclase prot	1.53e+02
32	87	5.6	169	1	W70228	Leishmania antigen Lmg	1.76e+02
33	87	5.6	193	1	P50343	EPO encoded by clone 1	1.76e+02
34	87	5.6	332	1	W12928	Mpl ligand analogue, [	1.76e+02
35	87	5.6	548	1	W88549	Secreted protein encod	1.76e+02
36	87	5.6	614	1	W20991	H. pylori inner membra	1.76e+02
37	87	5.6	2224	1	W04254	Human Factor V.	1.76e+02
38	87	5.6	2396	1	R29939	Deduced from Lelystad	1.76e+02
39	86	5.5	166	1	R71211	Human erythropoietin a	2.03e+02
40	86	5.5	345	1	W09405	Pineal gland specific	2.03e+02
41	86	5.5	368	1	W32868	Mitogen activated prot	2.03e+02
42	86	5.5	393	1	W32867	Mitogen activated prot	2.03e+02
43	86	5.5	746	1	R93270	GST-INL fusion interna	2.03e+02
44	86	5.5	2254	1	R76949	ACCase.	2.03e+02
45	86	5.5	2482	1	W23996	Human mitosin amino ac	2.03e+02

ALIGNMENTS

```
RESULT    1
ID   W37723 standard; Protein; 224 AA.
AC   W37723;
DT   09-JUN-1998  (first entry)
DE   Rat Hypertension related calcium regulator.
KW   Hypertension related calcium regulated gene; HCaRG; rat parathyroid;
KW   extracellular calcium concentration; antibody; hypertension;
KW   hyperthyroidism; osteoporosis; heart failure; diabetes; stroke;
KW   cancer; inflammatory disease; asthma.
OS   Rattus rattus.
FH   Key             Location/Qualifiers
FT   Misc_difference 15. .21
FT                   /note= "EF-hand like motif"
PN   WO9749807-A2.
PD   31-DEC-1997.
PF   23-JUN-1997; CA0439.
PR   21-JUN-1996; US-667495.
PA   (GOSS/) GOSSARD F.
PA   (HAME/) HAMET P.
PA   (LEWA/) LEWANCZUK R.
PA   (TREM/) TREMBLAY J.
PI   Gossard F, Hamet P, Lewanczuk R, Tremblay J;
DR   WPI; 98-077171/07.
DR   N-PSDB; V18890.
PT   Hypertension related calcium regulated gene - useful to develop
PT   products to treat or detect, e.g. hypertension, stroke,
PT   osteoporosis, heart failure, cancer, diabetes or asthma
PS   Claim 8; Pages 26-27; 46pp; English.
CC   This is the amino acid sequence of the hypertension related calcium
CC   regulated gene (HCaRG), which was isolated from the rat parathyroid.
CC   Its expression is regulated by extracellular calcium concentration.
CC   An antibody against the protein, can be used to detect or modulate
CC   (e.g. enhance or inhibit) abnormal calcium levels.  They can
CC   specifically be used to detect or treat, e.g. hypertension,
CC   hyperthyroidism, osteoporosis, heart failure, diabetes, stroke, cancer,
CC   inflammatory disease, and asthma.
SQ   Sequence   224 AA;

  Query Match          100.0%;  Score 1552;  DB 1;  Length 224;
  Best Local Similarity 100.0%;  Pred. No. 2.90e-120;
  Matches   224;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db       1 MSALGAAAPYLHHPADSHSGRVSFLGSQPSPEVTAVAQLLKDLDRSTFRKLLKLVVGALH 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       1 MSALGAAAPYLHHPADSHSGRVSFLGSQPSPEVTAVAQLLKDLDRSTFRKLLKLVVGALH 60

Db      61 GKDCREAVEQLGASANLSEERLAVLLAGTHTLLQQALRLPPASLKPDAFQEELQELGIPQ 120
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      61 GKDCREAVEQLGASANLSEERLAVLLAGTHTLLQQALRLPPASLKPDAFQEELQELGIPQ 120

Db     121 DLIGDLASLAFGSQRPLLDSVAQQQGSSLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQL 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     121 DLIGDLASLAFGSQRPLLDSVAQQQGSSLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQL 180

Db     181 KLTDGSAHRFEVPIAKFQELRYSVALVLKEMAELEKKCERKLQD 224
          |||||||||||||||||||||||||||||||||||||||||||||
Qy     181 KLTDGSAHRFEVPIAKFQELRYSVALVLKEMAELEKKCERKLQD 224


RESULT    2
ID   W88390 standard; Protein; 169 AA.
AC   W88390;
DT   26-APR-1999  (first entry)
DE   Human Zneul partial polypeptide.
KW   Zneu-1; neuro-growth factor-like protein; human; breast cancer;
KW   glioblastoma; pituitary adenoma; Alzheimer's disease; therapy;
KW   nerve regeneration; haematopoiesis; fertility; contraception;
KW   antibody.
OS   Homo sapiens.
PN   WO9857983-A2.
PD   23-DEC-1998.
PF   18-JUN-1998; U12763.
PR   18-JUN-1997; US-878322.
PR   18-JUN-1997; US-050143.
PA   (ZYMO ) ZYMOGENETICS INC.
PI   Blumberg H, Jelinek LJ, Lehner JM, Sheppard PO,
PI   Whitmore TE;
DR   WPI; 99-095324/08.
PT   New mammalian Zneul polypeptides - used to, e.g. treat Alzheimer's
PT   disease, cancer and to repopulate blood cells
PS   Claim 6; Page 57-58; 70pp; English.
CC   This claimed polypeptide comprises amino acid residues 105-273 of
CC   the human Zneul mature protein (see W88382), i.e. the 2 epidermal
CC   growth factor-like domains (see also W88384-85) and the second
CC   HSMHC3W5A homology domain of Zneul.  Zneul is a new neuro-growth
CC   factor-like protein that can be used as a growth, maintenance, or
CC   differentiation factor in the spinal cord, heart, spleen, testis,
CC   thyroid and lymph nodes.  Zneul may also play a role in breast
CC   cancer, glioblastomas, and pituitary adenomas.  Zneul can be used
CC   to treat Alzheimer's disease, cancer, to repopulate blood cells
CC   after chemotherapy, to stimulate myofibroblast proliferation,
CC   stimulate or inhibit growth factors made in the placenta, in
CC   fertility and contraception, or to regenerate nerves.  Claimed
CC   Zneul polypeptides (see also W88382-97), including specific domains
CC   of Zneul and epitope-bearing portions of Zneul, can be used to
CC   raise specific antibodies for use e.g. in diagnostic assays.
SQ   Sequence   169 AA;

  Query Match             7.7%;  Score 119;  DB 1;  Length 169;
  Best Local Similarity  42.0%;  Pred. No. 1.44e+00;
  Matches    21;  Conservative   13;  Mismatches  13;  Indels   3;  Gaps   3;

Db      93 KEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLLVHSFQQ 142
           :| |::| :  :| ||:| ::||  |:|  ||| : || |:||   :||:
Qy      65 REAVEQLGASANLSEERLAVLLAGTHTLLQQAL-R-LP-PASLKPDAFQE 111


RESULT    3
ID   W88391 standard; Protein; 181 AA.
AC   W88391;
DT   26-APR-1999  (first entry)
DE   Human Zneul partial polypeptide.
KW   Zneu-1; neuro-growth factor-like protein; human; breast cancer;
KW   glioblastoma; pituitary adenoma; Alzheimer's disease; therapy;
KW   nerve regeneration; haematopoiesis; fertility; contraception;
KW   antibody.
OS   Homo sapiens.
PN   WO9857983-A2.
PD   23-DEC-1998.
PF   18-JUN-1998; U12763.
PR   18-JUN-1997; US-878322.
PR   18-JUN-1997; US-050143.
PA   (ZYMO ) ZYMOGENETICS INC.
PI   Blumberg H, Jelinek LJ, Lehner JM, Sheppard PO,
PI   Whitmore TE;
DR   WPI; 99-095324/08.
PT   New mammalian Zneul polypeptides - used to, e.g. treat Alzheimer's
PT   disease, cancer and to repopulate blood cells
PS   Claim 6; Page 58; 70pp; English.
CC   This claimed polypeptide comprises amino acid residues 1-85 and
CC   159-254 of the human Zneul mature protein (see W88382), i.e. the
CC   first HSMHC3W5a homology domain fused to the second HSMHC3W5a
CC   homology domain of Zneul.  Zneul is a new neuro-growth factor-like
CC   protein that can be used as a growth, maintenance, or
CC   differentiation factor in the spinal cord, heart, spleen, testis,
CC   thyroid and lymph nodes.  Zneul may also play a role in breast
CC   cancer, glioblastomas, and pituitary adenomas.  Zneul can be used
CC   to treat Alzheimer's disease, cancer, to repopulate blood cells
CC   after chemotherapy, to stimulate myofibroblast proliferation,
CC   stimulate or inhibit growth factors made in the placenta, in
CC   fertility and contraception, or to regenerate nerves.  Claimed
CC   Zneul polypeptides (see also W88382-97), including specific domains
CC   of Zneul and epitope-bearing portions of Zneul, can be used to
CC   raise specific antibodies for use e.g. in diagnostic assays.
SQ   Sequence   181 AA;

  Query Match             7.7%;  Score 119;  DB 1;  Length 181;
  Best Local Similarity  42.0%;  Pred. No. 1.44e+00;
  Matches    21;  Conservative   13;  Mismatches  13;  Indels   3;  Gaps   3;

Db     105 KEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLLVHSFQQ 154
           :| |::| :  :| ||:| ::||  |:|  ||| : || |:||   :||:
Qy      65 REAVEQLGASANLSEERLAVLLAGTHTLLQQAL-R-LP-PASLKPDAFQE 111


RESULT    4
ID   W88382 standard; Protein; 254 AA.
AC   W88382;
DT   26-APR-1999  (first entry)
DE   Human neuro-growth factor-like protein Zneul mature polypeptide.
KW   Zneu-1; neuro-growth factor-like protein; human; breast cancer;
KW   glioblastoma; pituitary adenoma; Alzheimer's disease; therapy;
KW   nerve regeneration; haematopoiesis; fertility; contraception;
KW   antibody.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Domain          1. .85
FT                   /note= "hydrophilic domain (HSM1), homologous to
FT                         an HSMHC3W5A domain "
FT   Domain          86. .116
FT                   /note= "epidermal growth factor-like domain 1"
FT   Domain          117. .158
FT                   /note= "epidermal growth factor-like domain 2"
FT   Domain          159. .254
FT                   /note= "domain HSM2 homologous to an HSMHC3W5A
FT                         domain"
PN   WO9857983-A2.
PD   23-DEC-1998.
PF   18-JUN-1998; U12763.
PR   18-JUN-1997; US-878322.
PR   18-JUN-1997; US-050143.
PA   (ZYMO ) ZYMOGENETICS INC.
PI   Blumberg H, Jelinek LJ, Lehner JM, Sheppard PO,
PI   Whitmore TE;
DR   WPI; 99-095324/08.
PT   New mammalian Zneul polypeptides - used to, e.g. treat Alzheimer's
PT   disease, cancer and to repopulate blood cells
PS   Claim 6; Page 48-49; 70pp; English.
CC   This polypeptide comprises human Zneul mature polypeptide.  Zneul
CC   is a new neuro-growth factor-like protein (see also W88381).  Its
CC   closest human homologue is HSMHC3W5A, a gene in the HLA class III
CC   region, which is contained in a cosmid which contains Notch 4.
CC   Zneul is also homologous to Notch 4 in its EGF-like domains and may
```

```
CC   be involved in EGF receptor pathways. Zneu1 is widely expressed in
CC   adult tissues, with high expression in heart, placenta, spleen,
CC   testis, thyroid, spinal cord and lymph node. Zneu1 can be used as
CC   a growth, maintenance, or differentiation factor in the spinal
CC   cord, heart, spleen, testis, thyroid and lymph nodes. It may
CC   play a role in breast cancer, glioblastomas, and pituitary
CC   adenomas. Zneu1 may also be used to treat Alzheimer's disease,
CC   cancer, to repopulate blood cells after chemotherapy, to stimulate
CC   myofibroblast proliferation, stimulate or inhibit growth factors
CC   made in the placenta, in fertility and contraception, or to
CC   regenerate nerves. Claimed Zneu1 polypeptides (see also W88382-97),
CC   including specific domains of Zneu1 and epitope-bearing portions of
CC   Zneu1, can be used to raise specific antibodies for use e.g. in
CC   diagnostic assays.
SQ   Sequence   254 AA;

  Query Match            7.7%;  Score 119;  DB 1;  Length 254;
  Best Local Similarity  42.0%;  Pred. No. 1.44e+00;
  Matches    21;  Conservative   13;  Mismatches  13;  Indels   3;  Gaps   3;

Db    178 KEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLLVHSFQQ 227
          :| |::| :  :| ||:| ::||  |:|  ||| : || |:||   :||:
Qy     65 REAVEQLGASANLSEERLAVLLAGTHTLLQQAL-R-LP-PASLKPDAFQE 111


RESULT    5
ID   W88381 standard; Protein; 273 AA.
AC   W88381;
DT   26-APR-1999  (first entry)
DE   Human neuro-growth factor-like protein Zneu1.
KW   Zneu-1; neuro-growth factor-like protein; human; breast cancer;
KW   glioblastoma; pituitary adenoma; Alzheimer's disease; therapy;
KW   nerve regeneration; haematopoiesis; fertility; contraception;
KW   antibody.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Peptide         1. .19
FT                   /note= "putative signal peptide"
FT   Protein         20. .273
FT                   /note= "mature protein"
FT   Domain          20. .104
FT                   /note= "hydrophilic domain (HSM1), homologous to
FT                         an HSMHC3W5A domain "
FT   Domain          105. .135
FT                   /note= "epidermal growth factor-like domain 1"
FT   Domain          136. .177
FT                   /note= "epidermal growth factor-like domain 2"
FT   Domain          178. .273
FT                   /note= "domain HSM2 homologous to an HSMHC3W5A
FT                         domain"
PN   WO9857983-A2.
PD   23-DEC-1998.
PF   18-JUN-1998; U12763.
PR   18-JUN-1997; US-878322.
PR   18-JUN-1997; US-050143.
PA   (ZYMO ) ZYMOGENETICS INC.
PI   Blumberg H, Jelinek LJ, Lehner JM, Sheppard PO,
PI   Whitmore TE;
DR   WPI; 99-095324/08.
DR   N-PSDB; V84341.
PT   New mammalian Zneu1 polypeptides - used to, e.g. treat Alzheimer's
PT   disease, cancer and to repopulate blood cells
PS   Claim 6; Page 47-48; 70pp; English.
CC   This polypeptide comprises human Zneu1, a new neuro-growth factor-
CC   like protein. Its amino acid sequence was deduced from the
CC   nucleotide sequence (see V84341) of a cDNA clone isolated from a
CC   brain cDNA library. Zneu1's closest human homologue is HSMHC3W5A,
CC   a gene in the HLA class III region, which is contained in a cosmid
CC   which contains Notch 4. Zneu1 is also homologous to Notch 4 in its
CC   EGF-like domains and may be involved in EGF receptor pathways.
CC   Zneu1 is widely expressed in adult tissues, with high expression in
CC   heart, placenta, spleen, testis, thyroid, spinal cord and lymph
CC   node. Zneu1 polypeptide can be used as a growth, maintenance, or
CC   differentiation factor in the spinal cord, heart, spleen, testis,
CC   thyroid and lymph nodes. It may also play a role in breast cancer,
CC   glioblastomas, and pituitary adenomas. Zneu1 may be used to treat
CC   Alzheimer's disease, cancer, to repopulate blood cells after
CC   chemotherapy, to stimulate myofibroblast proliferation, stimulate
CC   or inhibit growth factors made in the placenta, in fertility and
CC   contraception, or to regenerate nerves. Claimed Zneu1
CC   polypeptides (see also W88382-97), including specific domains of
CC   Zneu1 and epitope-bearing portions of Zneu1, can be used to raise
CC   specific antibodies for use e.g. in diagnostic assays.
SQ   Sequence   273 AA;

  Query Match            7.7%;  Score 119;  DB 1;  Length 273;
  Best Local Similarity  42.0%;  Pred. No. 1.44e+00;
  Matches    21;  Conservative   13;  Mismatches  13;  Indels   3;  Gaps   3;

Db    197 KEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLLVHSFQQ 246
          :| |::| :  :| ||:| ::||  |:|  ||| : || |:||   :||:
Qy     65 REAVEQLGASANLSEERLAVLLAGTHTLLQQAL-R-LP-PASLKPDAFQE 111


RESULT    6
ID   W71684 standard; Protein; 195 AA.
AC   W71684;
DT   04-DEC-1998  (first entry)
DE   Amino acid sequence of the human tumourigenesis associated protein.
KW   Human; tumourigenesis associated protein; HTAP; transplantation;
KW   tumour; Antagonist; cancer; inflammation; immunological disease;
KW   antibody; probe; primer; PCR; amplification; hybridisation;
KW   inhibition.
OS   Homo sapiens.
PN   WO9841635-A1.
PD   24-SEP-1998.
PF   20-MAR-1998; U06066.
PR   20-MAR-1997; US-822260.
PA   (INCY-) INCYTE PHARM INC.
PI   Goli SK, Hillman JL;
DR   WPI; 98-521224/44.
DR   N-PSDB; V58281.
PT   New tumorigenesis-associated protein and related nucleic acid,
PT   vectors, transformed cells - antibodies, agonists and antagonists,
PT   for diagnosis, treatment and prevention of abnormal cellular
PT   differentiation, particularly cancers and inflammation
PS   Claim 1; Fig 1A-1B; 54pp; English.
CC   This is the amino acid sequence of the human tumourigenesis
CC   associated protein (HTAP), used in the method of the invention. HTAP,
CC   is involved with cell proliferation and inflammation. It can be used
CC   to stimulate cell proliferation (e.g. of cells intended for
CC   transplantation in treatment of tumours or infections, or to treat
CC   genetic defects). Antagonists of HTAP are used to treat or prevent
CC   a wide range of cancers (adenocarcinoma, melanoma, sarcoma, lymphoma,
CC   leukaemia etc.), also inflammation where associated with infection or
CC   immunological disease (e.g. asthma, cystic fibrosis, rheumatoid
CC   arthritis). HTAP is also used to raise antibodies and to screen
CC   libraries for specific-binding agents. The antibodies are used as for
CC   diagnosis or monitoring of HTAP-related diseases (in usual
CC   immunoassays), in competitive drug screens and to isolate HTAP from
CC   its natural sources. HTAP derived probes or primers, are used in
CC   standard amplification or hybridisation tests to diagnose HTAP-related
CC   diseases; to identify related sequences; for genomic mapping and for
CC   screening for specific inhibitors.
SQ   Sequence   195 AA;

  Query Match            7.6%;  Score 118;  DB 1;  Length 195;
  Best Local Similarity  23.2%;  Pred. No. 1.69e+00;
  Matches    22;  Conservative   29;  Mismatches  40;  Indels   4;  Gaps   4;

Db     98 ELFCTEYQNNKNSLEILLGSIGRSLPHITDVSWRLEYQIKTNQLHRMYRPAYLVTLSVQN 157
          :|    : : :  |: :    | ||||::   ||::  | |:   |  :|: |: | : :
Qy    125 DLASLAFGSQRPLLDSVAQQQGSSLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQLKLTD 184

Db    158 TDSPSYPEISFSCSMEQLQDLVG-KLKDASKSLER 191
            : : |::::   ::|:  |:  ||: :  ||:
```

```
Qy    185 GSAHRF-EVPIA-KFQELRYSVALVLKEMAE-LEK 216


RESULT    7
ID   Y07867 standard; Protein; 196 AA.
AC   Y07867;
DT   06-JUL-1999  (first entry)
DE   Human secreted protein fragment encoded from gene 16.
KW   Human; secreted protein; treatment; prevention; protein therapy; AIDS;
KW   gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
KW   developmental abnormality; fetal deficiency; blood disorder; leukemia;
KW   immune system disease; autoimmune disease; hepatic disease; lymphoma;
KW   renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;
KW   cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;
KW   pulmonary disorder; transplant rejection; osteoclast; osteoporosis;
KW   arthritis; malignancy; digestive; endocrine; infection.
OS   Homo sapiens.
PN   WO9918208-A1.
PD   15-APR-1999.
PF   01-OCT-1998; U20775.
PR   02-OCT-1997; US-060884.
PR   02-OCT-1997; US-060833.
PR   02-OCT-1997; US-060836.
PR   02-OCT-1997; US-060837.
PR   02-OCT-1997; US-060838.
PR   02-OCT-1997; US-060839.
PR   02-OCT-1997; US-060843.
PR   02-OCT-1997; US-060862.
PR   02-OCT-1997; US-060866.
PR   02-OCT-1997; US-060874.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Carter KC, Duan DR, Endress GA, Feng P, Ferrie AM,
PI   Florence KA, Greene JM, Janat F, Lafleur DW, Ni J,
PI   Rosen CA, Ruben SM, Shi Y, Young P, Yu G;
DR   WPI; 99-264022/22.
DR   N-PSDB; X37466.
PT   New isolated human genes and the secreted polypeptides they encode
PS   Claim 1b; Page 285-286; 368pp; English.
CC   This invention describes novel isolated human genes and the secreted
CC   proteins they encode. The products of the invention are useful for
CC   preventing, treating or ameliorating medical conditions, e.g. by protein
CC   or gene therapy. Also pathological conditions can be diagnosed by
CC   determining the amount of the new polypeptides in a sample or by
CC   determining the presence of mutations in the new polynucleotides.
CC   Specific uses are described for each of the 101 polynucleotides, based on
CC   which tissues they are most highly expressed in, and include developing
CC   products for the diagnosis or treatment of cancer, tumours,
CC   neurodegenerative disorders, developmental abnormalities and fetal
CC   deficiencies, blood disorders, leukemias, diseases of the immune system,
CC   autoimmune diseases, hepatic and renal disease, lymphomas, inflammation,
CC   allergies, Alzheimer's and cognitive disorders, schizophrenia, prostate
CC   disease, skeletal or cardiac muscle disorders, pulmonary disorders,
CC   transplant rejection, disorders involving osteoclasts such as
CC   osteoporosis, arthritis or malignancies, digestive/endocrine disorders,
CC   infections and AIDS. The human secreted proteins of the invention are
CC   represented in Y07852-Y07993 and the encoding nucleic acids are
CC   represented in X37451-X37552.
SQ   Sequence   196 AA;

  Query Match           7.6%;  Score 118;  DB 1;  Length 196;
  Best Local Similarity  23.2%;  Pred. No. 1.69e+00;
  Matches    22;  Conservative   29;  Mismatches  40;  Indels   4;  Gaps   4;

Db     98 ELFCTEYQNNKNSLEILLGSIGRSLPHITDVSWRLEYQIKTNQLHRMYRPAYLVTLSVQN 157
           :|    : : :  |: :    | ||||::   ||:: | |:   |  :|: |: | : :
Qy    125 DLASLAFGSQRPLLDSVAQQQGSSLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQLKLTD 184

Db    158 TDSPSYPEISFSCSMEQLQDLVG-KLKDASKSLER 191
          : : |::::   ::|:  |:  ||: :  ||:
Qy    185 GSAHRF-EVPIA-KFQELRYSVALVLKEMAE-LEK 216


RESULT    8
ID   R85889 standard; Protein; 603 AA.
AC   R85889;
DT   13-SEP-1996  (first entry)
DE   WD-40 domain-contg. rat insulin-like growth factor binding protein.
KW   WD40 repeat region; beta-transducin; protein-protein interaction; drug;
KW   intracellular signalling; protein kinase C; homology; motif; modulator;
KW   receptors of activated protein kinase; enzyme activity; isozyme; human.
OS   Rattus rattus.
PN   WO9521252-A2.
PD   10-AUG-1995.
PF   31-JAN-1995; U01210.
PR   01-FEB-1994; US-190802.
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PI   Mochly-Rosen D,  Ron D;
DR   WPI; 95-283772/37.
PT   New WD-40 (beta-transducin)-derived polypeptide(s) - which alter the
PT   activity of a protein, eg. protein kinase C, which interacts with a
PT   protein contg. a WD-40 region.
PS   Example 5; Page 125-128; 351pp; English.
CC   Proteins R85851-92 are protein which contain at least one WD-40 (also
CC   called beta-transducing homologous) amino acid repeat motifs. The WD-40
CC   regions are involved in protein-protein interactions between proteins
CC   involved in intracellular signalling.  An example of such an interaction
CC   is between protein kinase C and receptors of activated protein kinase
CC   (RACK), esp. RACK-1 (R85850).  Proteins R85851-82 were isolated based on
CC   homology with beta-transducin, whereas proteins R85882-92 were isolated
CC   based on homology with the WD-40 consensus sequence (R85893).  The
CC   proteins were used to construct the peptides R84928-R85063 and
CC   R85786-R85842.  The peptides can be used to identify target proteins
CC   contg. WD-40 motifs, as modulators of enzyme esp. isozyme, activity of
CC   proteins involved in protein-protein interaction and to screen for drugs
CC   that will affect protein-protein interaction involving WD-40 domains.
SQ   Sequence   603 AA;

  Query Match           6.7%;  Score 104;  DB 1;  Length 603;
  Best Local Similarity  31.9%;  Pred. No. 1.46e+01;
  Matches    38;  Conservative   31;  Mismatches  36;  Indels  14;  Gaps  13;

Db    272 DLSHNRVAGLMEDTFPGLLGLHVLRLAHNAIASLRPRTFKDLHFLEELQLGHNRIRQLGE 331
          :|| :|:| |:  |   ||  : |||:  : |||:| :| :   |:|| : :: | :|:
Qy     76 NLSEERLAVLLAGT-HTLLQ-QALRLP--P-ASLKPDAFQE-E-LQELGIPQDLIGDLAS 128

Db    332 RTFEGLGQLEVL-TLNDNQITEV-RVGAFSGLFNVAVMNLSGNCLRSLPERVFQGLDKL 388
           :| | :|  :| :: : | : : :|: |    :||: : |:  |||   |: | ||
Qy    129 LAF-G-SQRPLLDSVAQQQGSSLPHVSYFRWRVDVAI-STSAQS-RSLQPSVLMQL-KL 182


RESULT    9
ID   R42079 standard; Protein; 262 AA.
AC   R42079;
DT   29-APR-1994  (first entry)
DE   Impatiens Necrotic Spot Virus S nucleocapsid protein.
KW   INSV; tospovirus; resistance; pan-handle secondary structure;
KW   plant RNA virus.
OS   Impatiens Necrotic Spot Virus.
PN   EP-566525-A.
PD   20-OCT-1993.
PF   16-MAR-1993; 810190.
PR   19-MAR-1992; GB-006016.
PA   (SANO ) SANDOZ LTD.
PA   (VGRI/) VAN GRINSVEN M Q J M.
PA   (SANO ) SANDOZ PATENT GMBH.
PA   (SANO ) SANDOZ-ERFINDUNGEN VERW GES MBH.
PA   (SANO ) SANDOZ AG.
PI   De Haan PT,  Gielen JJL,  Goldbach RW,  Peters D,  Van Grinsven MQJ;
PI   Van Grinsven MQJM;
DR   WPI; 93-329814/42.
DR   N-PSDB; Q49956, Q49958.
PT   Recombinant INSV DNA constructs comprising INSV DNA coding for
PT   transcription into INSV RNA sequences - are used to transform
PT   plants to reduce susceptibility to INSV infection
PS   Example 4; Page 44-45; 64pp; English.
CC   INSV RNA was purified from systemically infected Nicotiana rustica
```

```
CC   leaves. The short (S) transcript comprises two genes, one coding for
CC   a non-structural protein and the other (on the complementary strand)
CC   coding for the nucleocapsid protein. The 5'- and 3'-terminal
CC   sequences of the S RNA are capable of hybridising to each other; the
CC   double-stranded structure obtained by such hybridisation is referred
CC   to as a "pan-handle".
SQ   Sequence   262 AA;

  Query Match            6.6%;  Score 103;  DB 1;  Length 262;
  Best Local Similarity  36.2%;  Pred. No. 1.69e+01;
  Matches    17;  Conservative   12;  Mismatches  17;  Indels   1;  Gaps   1;

Db    137 ATTDLKSCVLGVLLGGSLPLIASVLNFEIAALVPAIYQDAKHVELGI 183
          |:::|    |:|||:|:  |:   | :  |:| |  :|:  : ||||
Qy     73 ASANLSEERLAVLLAGTHTLLQQALRLPPASLKPDAFQEELQ-ELGI 118


RESULT   10
ID   R85888 standard; Protein; 605 AA.
AC   R85888;
DT   13-SEP-1996  (first entry)
DE   WD-40 domain-contg. insulin-like growth factor binding protein.
KW   WD40 repeat region; beta-transducin; protein-protein interaction; drug;
KW   intracellular signalling; protein kinase C; homology; motif; modulator;
KW   receptors of activated protein kinase; enzyme activity; isozyme; human.
OS   Synthetic.
PN   WO9521252-A2.
PD   10-AUG-1995.
PF   31-JAN-1995; U01210.
PR   01-FEB-1994; US-190802.
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
PI   Mochly-Rosen D,  Ron D;
DR   WPI; 95-283772/37.
PT   New WD-40 (beta-transducin)-derived polypeptide(s) - which alter the
PT   activity of a protein, eg. protein kinase C, which interacts with a
PT   protein contg. a WD-40 region.
PS   Example 5; Page 122-125; 351pp; English.
CC   Proteins R85851-92 are protein which contain at least one WD-40 (also
CC   called beta-transducing homologous) amino acid repeat motifs. The WD-40
CC   regions are involved in protein-protein interactions between proteins
CC   involved in intracellular signalling.  An example of such an interaction
CC   is between protein kinase C and receptors of activated protein kinase
CC   (RACK), esp. RACK-1 (R85850).  Proteins R85851-82 were isolated based on
CC   homology with beta-transducin, whereas proteins R85882-92 were isolated
CC   based on homology with the WD-40 consensus sequence (R85893).  The
CC   proteins were used to construct the peptides R84928-R85063 and
CC   R85786-R85842.  The peptides can be used to identify target proteins
CC   contg. WD-40 motifs, as modulators of enzyme esp. isozyme, activity of
CC   proteins involved in protein-protein interaction and to screen for drugs
CC   that will affect protein-protein interaction involving WD-40 domains.
SQ   Sequence   605 AA;

  Query Match            6.3%;  Score 98;  DB 1;  Length 605;
  Best Local Similarity  38.1%;  Pred. No. 3.57e+01;
  Matches    24;  Conservative   15;  Mismatches  17;  Indels   7;  Gaps   6;

Db    272 DLSHNRVAGLLEDTFPGLLGLRVLRLSHNAIASLRPRTFKDLHFLEELQLGHNRIRQLAE 331
          :|| :|:| ||  |   ||  : |||:  : |||:| :| :   |:|| : :: | :||
Qy     76 NLSEERLAVLLAGT-HTLLQ-QALRLP--P-ASLKPDAFQE-E-LQELGIPQDLIGDLAS 128

Db    332 RSF 334
           :|
Qy    129 LAF 131


RESULT   11
ID   W95014 standard; peptide; 178 AA.
AC   W95014;
DT   21-MAY-1999  (first entry)
DE   Chicken MGF sequence.
KW   Cytokine; interleukin-B30; IL-B30; forensic science; cell proliferation;
KW   inflammatory condition; drug screening; human; MGF.
OS   Gallus sp.
PN   WO9905280-A1.
PD   04-FEB-1999.
PF   24-JUL-1998; U15423.
PR   25-JUL-1997; US-900905.
PA   (SCHE ) SCHERING CORP.
PI   Bazan JF;
DR   WPI; 99-142935/12.
PT   Newly isolated or recombinant polynucleotide encoding mammalian
PT   cytokine interleukin-B30 (IL-B30), including fragments - useful for
PT   regulating activation, development, differentiation and function of
PT   various cell types, and for diagnosing and treating conditions
PT   associated with IL-B30
PS   Disclosure; Page 11-12; 83pp; English.
CC   This invention relates to mammalian cytokine interleukin-B30 (IL-B30)
CC   polypeptides. Host cells containing a vector comprising the IL-B30
CC   nucleic acids are used for the recombinant production of the proteins.
CC   The polynucleotides are useful for diagnosis of IL-B30 mediated
CC   conditions, and forensic science (e.g. to distinguish rodent from human,
CC   or as a marker to distinguish between different cells exhibiting
CC   differential expression or modification patterns). The IL-B30 (including
CC   fragments), together with antibodies that bind to IL-B30 are useful for
CC   teaching purposes. They are also used for treating conditions associated
CC   with abnormal physiology or development, including inflammatory
CC   conditions. The polypeptide cytokine should mediate cytokine synthesis
CC   and proliferation in cells. IL-B30 is useful for drug screening to
CC   identify compounds having binding affinity to IL-B30. The present
CC   sequence represents a chicken MGF.
SQ   Sequence   178 AA;

  Query Match            6.2%;  Score 97;  DB 1;  Length 178;
  Best Local Similarity  33.3%;  Pred. No. 4.14e+01;
  Matches    17;  Conservative   18;  Mismatches  13;  Indels   3;  Gaps   3;

Db     79 QIRAGLHAYHDSLGAVLRLLPNHTTLVETLQLDAANLSSN-IQQQMEDLGL 128
          |: |: :   : | ||| |   || | ::|:| :|:| :: :|::::::||:
Qy     70 QLGASANLSEERL-AVL-LAGTHTLLQQALRLPPASLKPDAFQEELQELGI 118


RESULT   12
ID   R36731 standard; Protein; 912 AA.
AC   R36731;
DT   26-AUG-1993  (first entry)
DE   Ubiquitin-specific protease, UBP3.
KW   Fusion protein; ubiquitin-methionine-beta-galactosidase; UBP2; UBP3;
KW   expression vector; ubiquitin-specific; protease; UBP1; in vitro;
KW   transformation; Ub-met-beta-Gal; in vivo.
OS   Synthetic.
PN   WO9309235-A.
PD   13-MAY-1993.
PF   06-NOV-1992; U09562.
PR   08-NOV-1991; US-789915.
PA   (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI   Baker RT,  Tobias JW,  Varshavsky A;
DR   WPI; 93-167700/20.
DR   N-PSDB; Q41290.
PT   Ubiquitin-specific protease(s) - specifically cleave fusion
PT   proteins regardless of size, at C-terminus of ubiquitin portion
PS   Claim 13; Page 58-64; 78pp; English.
CC   The sequences given in R36729-31 represent the ubiquitin-specific
CC   proteases UBP1, UBP2, and UBP3.  These proteases specifically cleave
CC   the 120 kD fusion protein ubiquitin-methionine-beta-galactosidase
CC   (Ub-met-beta-Gal) (see also R36728).  The proteases UBP1 and UBP2
CC   demonstrate activity both in vivo and in vitro, whereas UBP3 is only
CC   active in vivo.  These proteases have been shown to specifically
CC   cleave Ub-met-beta-Gal at the C-terminus of the ubiquitin moiety.
CC   The proteases can be used to deubiquinate fusion proteins in vivo,
CC   therefore prokaryotic cells having an expression vector one of these
CC   proteases can be further transformed with an expression vector
CC   encoding a ubiquitin fusion protein.  Such cells will then produce
CC   a deubiquinated product having a predetermined N-terminal amino acid
CC   residue.
SQ   Sequence   912 AA;
```

```
  Query Match             6.1%;  Score 95;  DB 1;  Length 912;
  Best Local Similarity  31.1%;  Pred. No. 5.56e+01;
  Matches    14;  Conservative   16;  Mismatches  15;  Indels   0;  Gaps   0;

Db    705 FRSVLDIPNNKESQSITLDPFQTIQLDISDAGVNDLETAFKKFSE 749
          ||  :|:: :  :|| :|:|   :|| ::|:: : :| :: || |
Qy    155 FRWRVDVAISTSAQSRSLQPSVLMQLKLTDGSAHRFEVPIAKFQE 199


RESULT   13
ID   W22601 standard; Protein; 4472 AA.
AC   W22601;
DT   27-FEB-1998  (first entry)
DE   Tylactone synthase ORF1 protein.
KW   Tylactone synthase gene cluster; tylG gene; multifunctional protein;
KW   polyketide; tylactone synthesis; antibiotic; tylosin.
OS   Streptomyces fradiae.
FH   Key             Location/Qualifiers
FT   Misc_difference 1
FT                   /note= "encoded by GTG"
FT   Domain          43. .447
FT                   /note= "ketosynthase domain, KSq(s)"
FT   Domain          586. .914
FT                   /note= "acyltransferase domain, AT(s)"
FT   Domain          954. .1038
FT                   /note= "acyl carrier protein domain, ACP(s)"
FT   Domain          1060. .1483
FT                   /note= "ketosynthase domain, KS1"
FT   Domain          1606. .1934
FT                   /note= "acyltransferase domain, AT1"
FT   Domain          2199. .2378
FT                   /note= "ketoreductase domain, KR1"
FT   Domain          2469. .2552
FT                   /note= "acyl carrier protein domain, ACP1"
FT   Domain          2576. .2999
FT                   /note= "ketosynthase domain, KS2"
FT   Domain          3149. .3477
FT                   /note= "acyltransferase domain, AT2"
FT   Domain          3502. .3687
FT                   /note= "dehydratase domain, DH2"
FT   Domain          4016. .4200
FT                   /note= "ketoreductase domain, KR2"
FT   Domain          4302. .4385
FT                   /note= "acyl carrier protein domain, ACP2"
PN   EP-791655-A2.
PD   27-AUG-1997.
PF   19-FEB-1997; 301056.
PR   22-FEB-1996; US-012078.
PA   (ELIL ) LILLY & CO ELI.
PI   Dehoff BS,  Kuhstoss SA,  Rosteck PR,  Sutton KL;
DR   WPI; 97-418046/39.
DR   N-PSDB; T80413.
PT   DNA encoding Streptomyces fradiae tylactone synthase domain - for
PT   production of tylosin-related polyketide compounds
PS   Claim 9; Pages 66-80; 220pp; English.
CC   W22601-W22605 represent proteins encoded by the tylactone synthase gene
CC   cluster of the invention. The gene cluster is also referred to as the
CC   tylG gene, and was isolated from Streptomyces fradiae. These sequences
CC   are multifunctional proteins which direct the synthesis of the polyketide
CC   tylactone, isolated from Streptomyces fradiae. Tylactone is the basic
CC   building block of the antibiotic tylosin. The DNA sequence can be
CC   modified so as to alter the type of carboxylic acids incorporated, the
CC   number of carboxylic acids incorporated and/or the post-condensation
CC   reactions performed, thereby resulting in novel tylosin-related
CC   polyketides.
SQ   Sequence   4472 AA;

  Query Match             6.1%;  Score 94;  DB 1;  Length 4472;
  Best Local Similarity  27.4%;  Pred. No. 6.43e+01;
  Matches    43;  Conservative   42;  Mismatches  59;  Indels  13;  Gaps  10;

Db   4287 GGAPAPAGGEGTAERLAGLGEEARLRELVRLVRAEVSGVLGYSGPDAVEPGRPFKDLGFD 4346
          |: |:|    : |: |  | : : :| |::||   |::: |    :|||   :  :|: :

Qy     26 GSQPSPEVT-AVAQLLKDL-DRSTFRKLLKLV---VGALHGKDCREAVEQLGASANLSEE 80

Db   4347 SLTAVELRNRLGAATGLRLPTALVFDRPTSQAVAEYLAAELAGPRD-GGDTAAAAFEGLE 4405
           |: :    :     :|||| |  : :|  :|  | |  ||: |:|  || |: || :
Qy     81 RLAVLLAGTHTLLQQALRLPPA-SL-KP--DAFQEELQ-ELGIPQDLIGDLASLAFGSQR 135

Db   4406 ALAAAVGALAEDDLRR-DVLRRRLTELAAALTPQGRN 4441
          :|  :|:      | :   :|:|: ::| : ::|:|:
Qy    136 PLLDSVAQQQGSSLPHVSYFRWRV-DVAISTSAQSRS 171


RESULT   14
ID   W05049 standard; Protein; 166 AA.
AC   W05049;
DT   10-DEC-1996  (first entry)
DE   Human thrombopoietin deletion variant dTPO-6.
KW   Human; thrombopoietin; TPO; variant; deletion mutant; megakaryocyte;
KW   growth factor; platelet production; thrombocytopoiesis.
OS   Synthetic.
PN   J08168386-A.
PD   02-JUL-1996.
PF   13-JUL-1995; 199299.
PR   19-OCT-1994; JP-278657.
PA   (SNOW ) SNOW BRAND MILK PROD CO LTD.
DR   WPI; 96-357243/36.
DR   N-PSDB; T41906.
PT   Thrombopoietin variants having megakaryocyte growth factor activity
PT   - useful for promoting growth of megakaryocytes and platelets, also
PT   as biochemical assay reagent
PS   Claim 13; Page 15; 21pp; Japanese.
CC   Four deletion variants of thrombopoietin (TPO), designated dTPO-5,
CC   -6, -7 and -8, were engineered. All four variants contained the
CC   first 131 residues from the N-terminal region of human TPO and all
CC   four variants promoted formation and growth of human megakaryocytes.
CC   Proteins comprising at least residues 1-131 of TPO are useful for
CC   promoting platelet production. The present sequence is that of dTPO-6
CC   which has the same sequence as wild-type TPO for the first 131
CC   amino acids, followed by residues 319-353 from the C-terminus of
CC   TPO.
SQ   Sequence   166 AA;

  Query Match             6.0%;  Score 93;  DB 1;  Length 166;
  Best Local Similarity  26.9%;  Pred. No. 7.44e+01;
  Matches    39;  Conservative   38;  Mismatches  56;  Indels  12;  Gaps  12;

Db     15 TARLTLSSPAPPACDLRVLSKLLRD-SHVLHSRLSQCPEVHPLPTPVLLPAVDFSLGEWK 73
          ::|::: :: |:: ::  :: ||:|  :    :|     | :|       ||:  ||
Qy     19 SGRVSFLGSQPSP-EVTAVAQLLKDLDRSTFRKLLK-LVVGALHGKDCREAVE-QLGA-S 74

Db     74 TQMEETKAQDIL-GAVTLLLEGVMAARGQLGPTCLSSLLGQLSGQVRLLLGALQSL-LGT 131
          : : | :   :| |: ||| :::  : : | |  :   | :| |  : |:| | || :|:
Qy     75 ANLSEERLAVLLAGTHTLLQQALRLPPASLKPDAFQEELQEL-GIPQDLIGDLASLAFGS 133

Db    132 QLHPLLPDPSAPTPTPTSPLLNTSY 156
          | :||| | |  :    |:| : ||
Qy    134 Q-RPLL-D-SV-AQQQGSSLPHVSY 154


RESULT   15
ID   R57140 standard; Protein; 405 AA.
AC   R57140;
DT   21-FEB-1995  (first entry)
DE   Mouse mucosal addressin cell adhesion molecule (MAdCAM-1).
KW   Mucosal addressin cell adhesion molecule; MAdCAM-1; leukocyte;
KW   endothelium; transmembrane protein; immunoglobulin.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   peptide         1. .21
FT                   /note= "signal peptide"
FT   region          362. .385
FT                   /note= "transmembrane region"
PN   WO9413312-A.
PD   23-JUN-1994.
```

PF 23-NOV-1993; U11404.
PR 15-DEC-1992; US-990866.
PA (STRD) UNIV LELAND STANFORD JUNIOR.
PI Briskin M, Butcher E;
DR WPI; 94-234206/28.
DR N-PSDB; Q69224.
PT DNA encoding a new mucosal addressin - is used to produce prods.
PT for diagnosis and therapy, e.g. modulating homing of leukocytes
PT to endothelial cells
PS Disclosure; Page 21; 30pp; English.
CC This mucosal addressin (MAd) protein is characterized by having
CC 3 Ig-like domains, a mucine-like region between the 2nd and 3rd
CC domains, a transmembrane domain and a cytoplasmic domain, where the
CC MAd is capable of binding to a leukocyte surface membrane protein.
CC The MAd cell adhesion molecule (MAdCAM-1) protein encoded by this
CC cDNA can be used for screening for molecules capable of binding to
CC MAdCAM-1 or inhibiting binding of leukocytes to MAdCAM-1. The
CC protein may be used to treat intestinal inflammatory diseases and
CC autoimmune diseases, and especially intestinal carcinomas, etc.
SQ Sequence 405 AA;

Query Match 6.0%; Score 93; DB 1; Length 405;
Best Local Similarity 31.6%; Pred. No. 7.44e+01;
Matches 18; Conservative 20; Mismatches 13; Indels 6; Gaps 5;

```
Db    168 EEIQEAEGTPLFRMTQ-RWRLP-SLGTPAPP-ALHCQVTMQLPKLV--LTHRKEIPV 219
          : : : :|::| :::  |||:  :::|:| : :|:  | ||| ||    :|| |:|:
Qy    139 DSVAQQQGSSLPHVSYFRWRVDVAISTSAQSRSLQPSVLMQL-KLTDGSAHRFEVPI 194
```

Search completed: Fri May 12 10:27:49 2000
Job time : 35 secs.